Manuscript prepared for Atmos. Meas. Tech.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 20 March 2016

Ground-based assessment of the bias and long-term stability of fourteen limb and occultation ozone profile data records

D. Hubert[1], J.-C. Lambert[1], T. Verhoelst[1], J. Granville[1], A. Keppens[1], J.-L. Baray[2,3], U. Cortesi[4], D. A. Degenstein[5], L. Froidevaux[6], S. Godin-Beekmann[7], K. W. Hoppel[8], E. Kyrölä[9], T. Leblanc[10], G. Lichtenberg[11], C. T. McElroy[12], D. Murtagh[13], H. Nakane[14,15], R. Querel[16], J. M. Russell III[17], J. Salvador[18], H. G. J. Smit[19], K. Stebel[20], W. Steinbrecht[21], K. B. Strawbridge[22], R. Stübi[23], D. P. J. Swart[24], G. Taha[25,26], A. M. Thompson[26], J. Urban[13,†], J. A. E. van Gijsel[27], P. von der Gathen[28], K. A. Walker[29,30], E. Wolfram[18], and J. M. Zawodny[31]

[1]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Laboratoire de l'Atmosphère et des Cyclones (Université de La Réunion, CNRS, Météo-France), OSU-Réunion (Université de la Réunion, CNRS), La Réunion, France
[3]Laboratoire de Météorologie Physique, Observatoire de Physique du Globe de Clermont-Ferrand (Université Blaise Pascal, CNRS), Clermont-Ferrand, France
[4]Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche, Sesto Fiorentino, Italy
[5]Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, SK, Canada
[6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[7]Laboratoire Atmosphère Milieux Observations Spatiales, Université de Versailles Saint-Quentin en Yvelines, Centre National de la Recherche Scientifique, Paris, France
[8]Naval Research Lab, Washington, DC, USA
[9]Finnish Meteorological Institute, Helsinki, Finland
[10]Jet Propulsion Laboratory, California Institute of Technology, Wrightwood, CA, USA
[11]German Aerospace Center (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, Germany
[12]York University, Toronto, ON, Canada
[13]Department of Earth and Space Sciences, Chalmers University of Technology, Göteborg, Sweden
[14]Kochi University of Technology, Kochi, Japan
[15]National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan
[16]National Institute of Water and Atmospheric Research, Lauder, New Zealand
[17]Department of Atmospheric and Planetary Science, Hampton University, VA, USA
[18]CEILAP-UNIDEF (MINDEF-CONICET), UMI-IFAECI-CNRS-3351, Villa Martelli, Argentina
[19]Research Centre Jülich, Institute for Energy and Climate Research: Troposphere (IEK-8), Jülich, Germany
[20]Norwegian Air Research Institute (NILU), Kjeller, Norway
[21]Meteorologisches Observatorium, Deutscher Wetterdienst, Hohenpeissenberg, Germany
[22]Air Quality Processes Research Section, Environment Canada, Toronto, ON, Canada
[23]Payerne Aerological Station, MeteoSwiss, Payerne, Switzerland
[24]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
[25]Universities Space Research Association, Greenbelt, MD, USA
[26]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[27]Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands
[28]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany
[29]Department of Physics, University of Toronto, Toronto, ON, Canada
[30]Department of Chemistry, University of Waterloo, Waterloo, ON, Canada
[31]NASA Langley Research Center, Hampton, VA, USA
[†]deceased

Correspondence to: D. Hubert (daan.hubert@aeronomie.be)

Abstract. The ozone profile records of a large number of limb and occultation satellite instruments are widely used to address several key questions in ozone research. Further progress in some domains depends on a more detailed understanding of these data sets, especially of their long-term stability and their mutual consistency. To this end, we made a systematic assessment of fourteen limb and occultation sounders that, together, provide more than three decades of global ozone profile measurements. In particular, we considered the latest operational Level-2 records by SAGE II, SAGE III, HALOE, UARS MLS, Aura MLS, POAM II, POAM III, OSIRIS, SMR, GOMOS, MIPAS, SCIAMACHY, ACE-FTS and MAESTRO. Central to our work is a consistent and robust analysis of the comparisons against the ground-based ozonesonde and stratospheric ozone lidar networks. It allowed us to investigate, from the troposphere up to the stratopause, the following main aspects of satellite data quality: long-term stability, overall bias, and short-term variability, together with their dependence on geophysical parameters and profile representation. In addition, it permitted us to quantify the overall consistency between the ozone profilers. Generally, we found that between 20–40 km the satellite ozone measurement biases are smaller than ±5 %, the short-term variabilities are less than 5–12 % and the drifts are at most ±5 % decade^{-1} (or even ±3 % decade^{-1} for a few records). The agreement with ground-based data degrades somewhat towards the stratopause and especially towards the tropopause where natural variability and low ozone abundances impede a more precise analysis. In part of the stratosphere a few records deviate from the preceding general conclusions; we identified biases of 10 % and more (POAM II and SCIAMACHY), markedly higher single-profile variability (SMR and SCIAMACHY), and significant long-term drifts (SCIAMACHY, OSIRIS, HALOE, and possibly GOMOS and SMR as well). Furthermore, we reflected on the repercussions of our findings for the construction, analysis and interpretation of merged data records. Most notably, the discrepancies between several recent ozone profile trend assessments can be mostly explained by instrumental drift. This clearly demonstrates the need for systematic comprehensive multi-instrument comparison analyses.

1 Introduction

Long-term global observations of the distribution and evolution of ozone are vital to improve our current understanding of atmospheric processes, and thereby to allow more robust projections of the recovery of the ozone layer and climate change. Measurements of the vertical profile of ozone have been carried out over the last few decades by a large number of instruments, operating in-situ or from remote vantage points, on the ground and in space (for an overview, see Hassler et al., 2014). These indisputibly show globally declining ozone levels during the 1980s and a large part of the 1990s in the lower and upper stratosphere (∼ 5–7 % decade^{-1}), and to a lesser extent also in the middle stratosphere (1–2 % decade^{-1}) (WMO, 2014; Harris et al., 2015). Furthermore, the observed loss rates are in excellent agreement with expectations for the chemical destruction of ozone by man-made halocarbons (WMO, 2014). The abundances of these substances have decreased significantly over the past 15–20 years (WMO, 2011), as a result of the Montreal Protocol and its subsequent adjustments and amendments. It is therefore generally expected that the ozone layer is currently recovering from the effects of ozone depleting substances , albeit in an atmosphere with concomitant increases in greenhouse gas concentrations and changes in residual circulation (Waugh et al., 2009; Oman et al., 2010). While observations provide substantial evidence for the levelling off of the downward trend around 1997 at most latitudes and altitudes, i.e. the first phase of recovery, it is less clear whether they support an upward trend in recent years (Harris et al., 2015). Whether the onset of the second stage has been detected (or not) is one of the key questions in current ozone research, a debate that is hampered by two factors. First, the small magnitude of the increases in ozone (a few percent) when compared to its natural variability. This can only be remedied by longer time series. And second, the lack of appropriate knowledge of the uncertainties in the observational records. Shedding more light on the latter issue is the main objective of this paper.

Limb and occultation sounders are of prime interest for ozone profile trend assessments, as they provide near-global coverage at reasonably high vertical resolution. However, satellite instruments are rarely operational for much more than a decade, so their records are generally combined for long-term studies. The uncertainties (overall bias, short-term variability and long-term stability) in the resulting combined data set are an intricate combination of the uncertainties inherited from the contributing data sets and those introduced by the merging algorithm. Tummon et al. (2015) recently noted that the former source of error tends to dominate over the latter, thereby demonstrating the need for a detailed characterization of each individual record and especially of their mutual consistency.

Numerous validation studies have been published in recent years (for an overview, see Lambert et al., 2016), but some important gaps remain. First of all, there are no comprehensive multi-instrument assessments of most limb/occultation sounders using ground-based data as a reference. Also satellite intercomparison studies rarely cover more than a handful of records (exceptions are e.g. Dupuy et al., 2009; Jones et al., 2009; Laeng et al., 2014). Tegtmeier et al. (2013) conducted perhaps the most complete assessment so far, of the

ozone climatologies from 18 sounders. Like most works, it was dedicated to the quantification of bias patterns and shorter-term variability, but not to a detailed assessment of the stability on decadal time scales. However, precise estimates of instrumental drift are crucial for a sound determination of the significance of trend results. Just a few (in some cases indirect) drift estimates are available from ground-based comparisons (e.g. Terao and Logan, 2007; Nair et al., 2012) or from satellite intercomparisons (e.g. Jones et al., 2009; Mieruch et al., 2012; Adams et al., 2014; Eckert et al., 2014). Moreover, no works comprise all the records considered in the recent trend assessments, by e.g. the World Meteorological Organisation (WMO) (WMO, 2014) or within the SPARC/IO3C/IGACO-O3/NDACC (SI2N) initiative (for an overview, see Harris et al., 2015). Finally, the quality of auxiliary pressure and temperature profiles plays a role too, as it unavoidably affects the quality of ozone data when used to convert the ozone profiles to another vertical coordinate (altitude $\leftrightarrow$ pressure) or ozone unit (number density $\leftrightarrow$ volume mixing ratio), a common step in the merging process. At the moment, very little information on this latter aspect of data quality is available.

Our objective is to shed more light on these three missing pieces of information. We therefore perform an exhaustive assessment, from the ground up to the stratopause, of the latest releases of the operational Level-2 ozone profile data sets collected by 14 limb/occultation instruments over the period 1984–2013: SAGE II (v7), SAGE III (v4), HALOE (v19), UARS MLS (v5), Aura MLS (v3.3), POAM II (v6), POAM III (v4), OSIRIS (v5.07), SMR (v2.1), GOMOS (IPF 6), MIPAS (ML2PP 6), SCIAMACHY (SGP 5), ACE-FTS (v3) and MAESTRO (v1.2). Each satellite data set is compared to the observations by the ground-based ozonesonde and stratospheric ozone lidar networks, thereby acting as a pseudo-global, independent and well-characterised transfer standard. The robust analysis of co-located satellite-ground profile pairs allows us to quantify overall bias, short-term variability and long-term stability of the satellite records, and their dependence on altitude, latitude and season. Methodology and results for the native profile representation of each record are described in Sects. 3–5. In Sect. 6 we investigate whether the accompanying ancillary meteorological data impact ozone data quality when the original profiles are converted to another vertical coordinate or ozone unit.

The adoption of a consistent analysis framework permits us to bring all single-instrument results together, and examine the mutual consistency between instruments of each quality indicator (Sect. 7). We report the tendencies and several peculiarities, most notably a few instruments that drift significantly at some altitudes. Finally, we frame our findings within the broader context (Sect. 8), by commenting on current challenges related to verifying user requirements, and by highlighting the implications of our results for the design of merging schemes. Perhaps the most tangible outcome of our study is the successful interpretation of discrepancies in recent trend studies in terms of instrumental drift. It demonstrates that our work can contribute to a better exploitation of the limb and occultation ozone profile data sets. This should, in the end, be beneficial not only for trend assessments and the related merging activities, but also for other applications, such as trend attribution studies or model evaluations.

2 Ozone profile data records

Our assessment covers the period between October 1984 and May 2013 and considers fourteen satellite missions and two types of ground-based instruments. We first present the ozone profile data records that play a central role in our analyses: those gathered by ozonesonde and stratospheric lidar instruments. Then, we introduce the limb and occultation sounders that are the subject of this work. We limit ourselves to brief descriptions since all space- and ground-based ozone profile measurement techniques were reviewed exhaustively by Hassler et al. (2014). The technical details most relevant to our assessment are summarised in Tables 1–3.

2.1 Ground-based network observations

2.1.1 Ozonesondes

Balloon-borne ozonesondes are launched around the world, at many sites at least once a week. These electrochemical instruments record ozone partial pressure in situ at high vertical resolution (100–150 m) from the surface to the middle stratosphere ($\sim$ 30–35 km). An interfaced radiosonde provides the pressure (p), temperature (T) and GPS data necessary to geolocate each measurement, or to convert ozone partial pressure to other units. The data quality depends on various factors such as sonde type and manufacturer, the preflight characterisation and postflight processing (see e.g. Tarasick et al., 2016; Van Malderen et al., 2016). However, when standard operating procedures are followed the three most commonly used sonde types[1] produce consistent results between the tropopause and $\sim 28\,\mathrm{km}$, with biases smaller than $\pm 5\,\%$ and precisions better than $\sim 3\,\%$ (Smit and ASOPOS-panel, 2014). At higher and lower altitudes the data quality degrades somewhat and the differences between the sonde types becomes more clear. Overall, ECC-type sondes perform best with a bias of $\pm$5–7 % and a precision of 3–5 % in the troposphere. We use the ozonesonde data acquired by the Network for the Detection of Atmospheric Composition Change (NDACC, http://www.ndacc.org), WMO's Global

[1]Nowadays more than 80 % of the stations launch an electrochemical concentration cell (ECC) sonde (Komhyr, 1969). The Brewer-Mast sonde has mostly been used by the early sounding stations with long data records (Brewer and Milford, 1960), while the Japanese stations fly a carbon iodine cell sonde (Kobayashi and Toyama, 1966).

Atmospheric Watch (GAW, data distributed by the World Ozone and Ultraviolet Data Centre http://www.woudc.org) and the Southern Hemisphere Additional Ozonesondes network (SHADOZ, http://croc.gsfc.nasa.gov/shadoz, Thompson et al. (2012)). The stations considered in this work are listed in Table 1, together with the total number of screened profiles over the analysis period. The screening procedure is outlined in Sect. 3.

2.1.2 Stratospheric ozone lidars

Differential absorption lidars are laser-based active remote sensing systems that operate mostly during clear-sky nights. Profiles of ozone number density versus geometric altitude are retrieved between the tropopause and 45–50 km from backscattered signals at two wavelengths (Mégie et al., 1977). While instrument and retrieval set-up differs from one site to another, the NDACC ozone lidar network can be considered homogeneous; biases between stations are less than 2 % between 20–35 km. In this altitude range both bias and precision are estimated at ~ 2 % and worsen to 5–10 % at other altitudes due to e.g. lower signal-to-noise ratios or the saturation of the detectors (Keckhut et al., 2004). The vertical resolution degrades for the same reason from 0.3 km around the tropopause to 3–5 km in the upper stratosphere (Godin et al., 1999). The retrieval algorithms used at the different sites were extensively intercompared and the profile measurements validated against a mobile lidar reference, ozonesondes and microwave radiometers (McGee et al., 1991; Keckhut et al., 2004). Comparisons to space-based observations over the range 20–40 km furthermore showed biases less than ± 5 % and a decadal stability better than ± 5 % decade^{-1} (Nair et al., 2012). We use data from thirteen stratospheric ozone lidars in the NDACC network. Geographical location, measurement period and number of screened profiles over the analysis period are listed in Table 2. The screening procedure is outlined in Sect. 3. Whenever lidar data are converted to non-native profile representations, we do so in this work using the p/T information extracted at the time and location of the lidar measurement from the ERA-Interim reanalysis fields (Dee et al., 2011) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).

2.2 Satellite observations

Over the past few decades numerous instruments were deployed in space to monitor atmospheric ozone. Detailed intercomparison studies of monthly zonal mean ozone profile data (i.e. Level-3) were recently published for nadir-viewing (Kramarova et al., 2013) and limb/occultation-viewing instruments (Tegtmeier et al., 2013). Here, we focus on a ground-based validation of the Level-2 ozone profile records from fourteen limb/occultation sounders that had (have) prime sensitivity in the stratosphere and were (are) operational for more than three years, see Table 3.

Most instruments were launched once: HALOE (Halogen Occultation Experiment), OSIRIS (Optical Spectrograph and InfraRed Imaging System), SMR (Sub-Millimetre Radiometer), GOMOS (Global Ozone Monitoring by Occultation of Stars), MIPAS (Michelson Interferometer for Passive Atmospheric Sounding), SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY), ACE-FTS (Atmospheric Chemistry Experiment Fourier Transform Spectrometer) and MAESTRO (Measurements of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation). Some were deployed more than once, with improved design: SAGE (Stratospheric Aerosol and Gas Experiment, II and III), MLS (Microwave Limb Sounder, on the UARS and EOS-Aura platforms) and POAM (Polar Ozone and Aerosol Measurement, II and III). Five instruments (OSIRIS, SMR, ACE-FTS, MAESTRO and Aura MLS) remain operational until the present, nine ceased operations before the end of the analysis period (May 2013).

We consider the latest data release of the operational Level-2 products (Table 3), which typically comprise not one but several data sets. Our focus is on the observations that are best suited for long-term studies of stratospheric ozone. We therefore choose the 205 GHz profiles for UARS MLS rather than the 183 GHz retrievals (Livesey et al., 2003). The standard Aura MLS product, considered here, is based on observations by the 240 GHz radiometer. We take the 501.8 GHz retrievals for SMR since these are less biased (although more noisy) than the 544.6 GHz data (Urban et al., 2005). MAESTRO retrievals in the visible range perform better in the upper stratosphere (US) than the ultraviolet product and are therefore used here (McElroy et al., 2007). For SAGE III, we consider the profiles retrieved with the multiple linear regression technique rather than the SAGE II type method used for v6.2 (Wang et al., 2006). We further select MIPAS data from the nominal measurement mode (70% of total number of observations) which is most suitable for long-term stratospheric studies (Raspollini et al., 2013). The ACE-FTS team provides ozone data sets on both a variable and a fixed altitude grid, we pick the latter product.

A number of alternative data sets for these instruments were not included in this assessment. For instance the retrievals by scientific prototype Level-2 processors (MIPAS, SCIAMACHY, ...) were not considered here. Their bias structure is often comparable to that of the operational ozone data set, especially when contrasted to that of other instruments (e.g. Rozanov et al., 2007; Laeng et al., 2015), due to the use of the same calibrated Level-1 radiance data and a common sensitivity to retrieval parameters (e.g. spectroscopic data). Profile data from alternative viewing geometries (e.g. lunar occultations for SAGE III and SCIAMACHY, solar occultation data from SCIAMACHY or bright limb measurements by GOMOS) were not investigated either, and

their quality may well be different from the findings presented in the following.

Table 3 summarises host platform, observation geometry and time, spectral region and spatial coverage. Vertical resolution and sampling in space and time are mainly determined by the observation geometry, the orbit and the spectral range. Solar occultation observations yield 30 profiles per day at $\sim 1\,\mathrm{km}$ vertical resolution. Limb instruments on the other hand easily provide 1000 profiles per day but with a poorer vertical resolution of $\sim 3\,\mathrm{km}$ and a larger uncertainty in the altitude registration as well. The latter often changes with time, e.g. the UARS MLS team noticed an upward drift of the geopotential height (GPH) of the $100\,\mathrm{hPa}$ reference level by $600\,\mathrm{m}$ between 1991 and 1997 (Livesey et al., 2003, Fig. 1). A downward drift of $100\,\mathrm{m}$ in GPH was also found in Aura MLS v3.3 data from 2005 to 2009, but stabilised thereafter (Livesey et al., 2015). Fortunately, the pressure information retrieved by limb emission instruments (including UARS and Aura MLS) is typically more reliable and therefore used as native vertical scale instead of altitude.

We screen the satellite profiles according to the prescriptions of the data provider (Table 3). In some cases this implies the removal of a considerable part of the data record, e.g. periods during which the product stability is not guaranteed. In particular, we remove the UARS MLS data after the 15 June 1997 switch-off of the $63\,\mathrm{GHz}$ radiometer (Livesey et al., 2003). We also reject MIPAS observations before January 2005 since these are potentially biased relative to those from the second phase of the mission due to a different set of retrieval microwindows (Ceccherini et al., 2013). Finally, from September 2010 onwards the ACE-FTS and MAESTRO retrievals are affected by problems with auxiliary input data and therefore rejected from the analysis. These issues were fixed in the v2.5/v3.5 data release of ACE-FTS, which extends the mission's record to the present. Data providers generally recommend a vertical range for their ozone product besides the standard screening prescriptions, see Table 3. Here, we keep all grid levels in order to verify at what point the data quality starts to degrade. Fig. 1 shows for each instrument the vertical range considered in this work.

Each record is provided in its native ozone profile representation (Fig. 1) defined by the vertical coordinate (altitude or pressure), the grid levels and the unit in which ozone is expressed (volume mixing ratio, VMR, or number density). The vertical grid of some records varies with the changing tangent heights of the measurements. Satellite data providers typically include the pressure and/or temperature data required to convert the native ozone profiles to another representation. These auxiliary data are sometimes retrieved by the same processor but in general taken from an external source (see Table 3). This assessment focuses primarily on data quality in the native representation. But given its importance in e.g. the data merging context we complement the analysis with tests of the impact of auxiliary data on the profile quality in other representations. We will see in Sect. 6 that this should indeed not be ignored. Throughout this work the difference between geometric altitude and geopotential height is neglected.

3 Analysis approach and data preprocessing

A careful design of the analysis allows us not only to obtain robust estimates of the data quality of the individual satellite records but also, and this is one of our primary objectives, to assess their mutual consistency. Prerequisite to achieving these goals is a good understanding of the metrological aspects of the comparison analysis. Our analysis approach is therefore based on three principles that reduce confounding methodological biases. First of all, we use a single analysis and software framework. Second, all satellite records are compared to the same reference data, from ground-based observations. And finally, the manipulation of satellite data is kept to a strict minimum. In this section we describe the general aspects of the analysis. A detailed account of how decadal stability, bias and short-term variability are estimated follows in Sects. 4 and 5.

The ozonesonde and lidar networks provide vertical ozone profiles of well-documented quality and serve as suitable transfer standards on a pseudo-global scale and from the troposphere to the stratopause. We compare the satellite profiles to co-located ground-based measurements in relative units

$$\Delta x_{ij}(l) = 100 \times \frac{x_{ij,\,\mathrm{sat}}(l) - x'_{ij,\,\mathrm{gnd}}(l)}{x'_{ij,\,\mathrm{gnd}}(l)}. \qquad (1)$$

Here, $x_{ij,\,\mathrm{sat}}(l)$ and $x'_{ij,\,\mathrm{gnd}}(l)$ represent respectively satellite and (vertically smoothed and representation-transformed) ground-based ozone at grid level l of co-location pair i for correlative instrument j. If satellite biases have a multiplicative nature then any time-dependence in ozone levels (e.g. seasonal, interannual, solar cycle) is divided out in the relative differences. Another advantage is that it allows a direct comparison between the results in different ozone units. A disadvantage, however, is that relative differences are sensitive to low ozone values, leading to larger values in and below the UTLS (upper troposphere lower stratosphere) and in the upper stratosphere.

The uncertainty on Δx is determined by several factors besides pure measurement and retrieval uncertainties ($\mathbf{S}_{x_{\mathrm{sat}}}$, $\mathbf{S}_{x_{\mathrm{gnd}}}$) because satellite and ground-based instruments have different perceptions of a variable atmosphere. Vertical and horizontal resolutions differ and the probed air masses rarely coincide perfectly in space and time. In addition, the comparison can only be done when both profiles are expressed in the same representation. As a result, the total comparison error budget contains terms related to the differences in smoothing, the spatio-temporal mismatch of the co-locations and the auxiliary data used to transform between profile representations. When correlations between the terms are disregarded, the total uncertainty covariance matrix (including systematic

and random components) becomes $\mathbf{S}_{\Delta x} = \mathbf{S}_{x_{\mathrm{sat}}} + \mathbf{S}_{x_{\mathrm{gnd}}} + \mathbf{S}_{\mathrm{smoothing}} + \mathbf{S}_{\mathrm{mismatch}} + \mathbf{S}_{\mathrm{auxiliary}}$ (von Clarmann, 2006). Furthermore, when Δx data are averaged or regressed the co-located profile sample may not be sufficiently representative of the actual state of the studied parameter (ozone differences). Toohey et al. (2013) recently showed the importance of $\mathbf{S}_{\mathrm{sampling}}$ for trace gas climatologies and Damadeo et al. (2014) for time series analyses. Estimating sampling uncertainty for validation purposes is an analysis in its own right and outside the scope of this paper.

The next few paragraphs describe the data preprocessing scheme in which the mitigation of the uncertainties due to differences in smoothing, geolocation and auxiliary data plays a central role. Preprocessing starts off by removing the unreliable measurements following the guidelines of the data providers. Table 3 lists the recommended screening procedure references for the satellite records. Ground-based data are filtered using general criteria, removing measurements with larger uncertainties: altitudes below 5 hPa (~ 33 km) for ozonesondes and outside the 15–47 km range for lidars. In addition, we reject measurement levels with clearly unphysical readings ($O_3 < 0$, $p < 0$ hPa, $T < 0$ K or $T > 400$ K) or during unrealistic jumps in pressure ($\mathrm{d}p/\mathrm{d}t > 0$ and $\mathrm{d}z > 0.1$ km). Entire profiles are discarded from further analysis when (a) more than half of the levels are tagged bad, or (b) less than 30 levels are tagged good.

The choice of a co-location window is a trade-off between mismatch uncertainties and a sufficiently large sample size to obtain robust statistical estimates. We found that a maximum horizontal distance Δr of 500 km between the profiles is optimal, given the typical horizontal resolution of the order of a few hundred km of the satellite and ground-based measurements. The maximal temporal separation Δt is 6 h for MIPAS and Aura MLS, and 12 h for the other instruments. When multiple satellite profiles are present in the co-location window around a ground-based profile only the pair closest in space and time is retained, defined by $\sqrt{\Delta r^2 + V_{\mathrm{wind}}^2 \Delta t^2}$ with $V_{\mathrm{wind}} = 100\,\mathrm{km\,h^{-1}}$ as a rough estimate of horizontal wind speed in the stratosphere. Multiple co-locations occur mostly between polar orbiting instruments and high latitude stations. Figure S1 in Supplement shows the latitude-time cross-section of the co-location samples.

Mismatch uncertainties $\mathbf{S}_{\mathrm{mismatch}}$ increase when and where atmospheric inhomogeneities are larger. Diurnal variations in ozone contribute a systematic component since the local time of ground-based observations (ozonesonde mostly around noon, lidar during night) and satellite measurements (Table 3) is generally constant. Biases due to the diurnal cycle are negligible below 30 km, but not at higher altitudes where ozone reaches minimal levels after dawn and maximal values in the afternoon (Schanz et al., 2014; Parrish et al., 2014; Sakazaki et al., 2015). The largest effect on our bias estimates is expected in the middle (< 2–3 %) and upper stratosphere (< 4 %) for the comparisons of lidar to sunset occultation profiles and, to a lesser extent, to the evening observations by SMR and OSIRIS. The random component of mismatch uncertainty is typically 5 % but can reach 20 % at e.g. Antarctic stations dropping in and out of the vortex (Cortesi et al., 2007; De Clercq, 2009).

There is no well-established method in the community to remove the horizontal component of the smoothing error. Instead we refer to the model-based estimates for the specific case of MIPAS comparisons (Cortesi et al., 2007; De Clercq, 2009), which indicated that the horizontal smoothing uncertainty mainly has a random nature and is of similar magnitude as the mismatch uncertainty. The vertical component on the other hand can be mostly removed by smoothing the ground-based profiles. We use a triangular response function with a base width that follows the altitude-dependent satellite resolution (Table 3). The exception is the MIPAS analysis, for which we smoothed with the vertical averaging kernel (AK) and a-priori of the co-located MIPAS profile. Such an AK smoothing was initially also tried for the SCIAMACHY analysis. Unfortunately it introduced peculiar and unexpected vertical oscillations in the comparisons, so we resorted to the triangular method for SCIAMACHY. The comparison results, especially observed spreads, differ slightly when another shape of the smoothing function is chosen (we tried rectangular and Gaussian windows), but most of the vertical smoothing error is removed. We estimate that the residual vertical smoothing uncertainty is less than a few percent.

In a final preprocessing step the data are transformed to the same profile representation, defined by the ozone unit (number density or VMR), the vertical coordinate (altitude or pressure) and the levels of the vertical grid. Differences between geometric and geopotential height are neglected. We focus on the satellite instrument's native representation, see Fig. 1, mainly because it is closest to the retrieved information, but also because users will use it as a starting point to convert to another representation if their application requires that. They can use the auxiliary pressure and/or temperature profiles provided along with the ozone profiles in many satellite records for this purpose. As we have seen in Sect. 2.2, these auxiliary data originate from different sources which may lead to a representation-dependence of the mutual consistency of the satellite data quality. This is discussed further in Sect. 6. Until then, only the correlative data are converted when needed. Ozonesonde data are transformed with the help of p and T measurements from the attached radiosonde, and lidar data using ERA-Interim fields. The quality of these ancillary data has been investigated by various authors (e.g. Sun et al. (2013); Stauffer et al. (2014); Simmons et al. (2014); Inai et al. (2015)). The regridding to the satellite's vertical grid is based on a pseudo-inverse interpolation method (Calisesi et al., 2005). Since the ground-based grid is more finely resolved than the satellite grid, the associated regridding uncertainties are generally negligible. We note that the SMR, GOMOS, MIPAS and MAESTRO profiles are inevitably regridded as well, since the grid is variable. In these cases the levels of the comparison grid are

selected to reflect the average spacing between two lines of sight.

To conclude this section we repeat the importance of using a single analysis and code framework. Apart from some unavoidable preprocessing steps, the data and analysis flow is identical for all fourteen satellite comparison studies. In this way, the methodological biases are mostly identical and, hence, unlikely responsible for eventually observed differences between the satellite records. This approach will be exploited in Sect. 7. The next two sections present a detailed assessment of the bias, the short-term variability and the decadal stability of each individual satellite record.

4 Decadal stability

We estimate the decadal stability of satellite data through a robust analysis of the time series of the satellite-ground differences. This is a two-step process, in which the linear drift is first estimated at each ground station and subsequently averaged over the ozonesonde and lidar networks. The focus of this section is on the decadal stability of the individual satellite records, in their native profile representation. Later on we expand the discussion to the consistency of drift between profile representations (Sect. 6) and between satellite records (Sect. 7).

4.1 Methodology

4.1.1 Time series analysis at individual stations

We first estimate the drift of the satellite data at each ground station. The comparison time series can contain large gaps and/or outliers, see e.g. the GOMOS comparisons in Fig. 2 (top panel). Hence robust techniques are needed to estimate not only the drift but also its uncertainty (Muhlbauer et al., 2009; Croux et al., 2004). To this end we use an iterative Tukey-bisquare reweighted least-squares procedure to fit the daily averaged relative difference time series to a linear regression model

$$\Delta x_{ij}(l) = \alpha_j(l)\,(t_i - t_0) + \beta_j(l) + e_{ij}(l). \quad (2)$$

With $\Delta x_{ij}(l)$ as in Eq. (1) at time t_i and grid level l, and the fit residual $e_{ij}(l)$. In this model, the fit parameter $\alpha_j(l)$ represents the linear drift of the satellite data relative to the ground-based record j, whereas $\beta_j(l)$ is the bias between both records at reference time t_0. Time series with less than ten data points are not regressed. The significance of the estimated $\hat{\alpha}(l)$ is tested using a robust estimate of its standard deviation $\hat{\sigma}_\alpha(l)$ proposed by Street et al. (1988), a slightly modified version of the ordinary least-squares expression. Figure 2 illustrates three time series with superimposed regression results (left panels, blue line) and the corresponding 95 % confidence intervals for $\hat{\alpha}(l)$ (right panels, vertical dashed blue lines).

4.1.2 Aggregation into ground network average

In a second step, the drift estimates $\hat{\alpha}_j(l)$ are averaged over various ground stations $j = \{1, \ldots, N\}$. None of the satellite records exhibit a clear latitudinal structure of drift (see for instance HALOE and Aura MLS at 25 km in Fig. 3). Therefore, we average the results over the entire sonde network and over the entire lidar network. Since there is a clear variability in the regression uncertainty across the network, each station estimate is weighted by the inverse of its variance $w_j(l) = \hat{\sigma}^{-2}_{\alpha,j}(l)$. The network-averaged drift

$$\bar{\alpha}(l) = \frac{\sum_j w_j(l)\,\hat{\alpha}_j(l)}{\sum_j w_j(l)}, \quad (3)$$

has a standard deviation $\sigma_{\bar{\alpha}}(l) = 1/\sqrt{\sum_j w_j(l)}$.

The single-site drift uncertainties alone do not always explain the observed variability of the drift estimates over the network. When the number of stations is large enough ($N \gtrsim 20$) the distribution of normalised residuals $\nu_j(l) = (\hat{\alpha}_j - \bar{\alpha})/\hat{\sigma}_{\alpha,j}$ should have unit variance for realistic estimates $\hat{\sigma}_\alpha$ of the variance of $\hat{\alpha}$. That is typically not the case for the dense samplers, that tend to have larger variance as illustrated for instance for Aura MLS in Fig. 3 (right). This suggests an unaccounted for source of uncertainty, likely related to differences in sampling or inhomogeneities across the ground-based network. We follow an ad-hoc approach to incorporate this unknown component, by scaling the uncertainty up

$$\sigma^*_{\bar{\alpha}}(l) = \kappa(l) \times \sigma_{\bar{\alpha}}(l), \quad (4)$$

so that the reduced $\chi^2(l) = \sqrt{\frac{1}{N-1}\sum_j \nu_j(l)^2}$ becomes unity. We also assume, conservatively, that the original regression uncertainty does not overestimate the true uncertainty, hence $\kappa(l) = \max\{\chi^2(l), 1\}$. In the following, this adjusted standard deviation $\sigma^*_{\bar{\alpha}}(l)$ is used to test the significance of the drift averages at the 5 % level. Figure S2 in Supplement shows the κ-adjustment factor for each satellite record.

4.1.3 Sensitivity to analysis parameters

The importance of correct single station uncertainties $\hat{\sigma}_{\alpha,j}(l)$ is evident for the calculation of both the weighted mean and its uncertainty. The possible presence of data gaps, outliers and auto-correlation in the time series led us to cross-check the analytic expression of Street et al. (1988) with a bootstrapping technique (Efron and Tibshirani, 1986). Each comparison time series was resampled 2500 times by replacement of single data points, and subsequently regressed to reconstruct the distribution of $\hat{\alpha}_j(l)$ (Fig. 2, right). The 2.5 % and 97.5 % quantiles define the 95 % confidence interval (light red area) which is in good agreement with the analytic expression (vertical dashed blue lines). Replacing the analytic by the bootstrap-derived uncertainties in Eq. (3) and (4)

changes $\bar{\alpha}$ and $\sigma^*_{\bar{\alpha}}$ typically by less than $\sim 0.5\,\%\,\mathrm{decade}^{-1}$ (Fig. S17 in Supplement). Figure 2 (left) also illustrates the outcome of other sensitivity checks, such as changing the temporal resolution of the time series prior to regression (from daily to monthly, green curve) or adding a 1-year harmonic component to the regression model (orange curve). Again, the results are very consistent, changing $\bar{\alpha}$ and $\sigma^*_{\bar{\alpha}}$ typically by less than $\sim 1\,\%\,\mathrm{decade}^{-1}$ (Fig. S17 in Supplement). These cross-checks demonstrate the robustness of the results obtained in the rest of this paper.

4.2 Selection of ground sites

Several ground sites were discarded from the drift analysis because of poor sampling, spurious features in the reference data (e.g. due to changes instrumentation, ...), or peculiarities in the satellite data. Drift estimates at stations with small co-location samples have large uncertainty and, hence, in principle a negligible influence on the network-averaged estimates. Nevertheless, a few ozonesonde stations with a short data record or with episodic observations collected during field campaigns are not retained for the regression analyses. Figure 4 shows the vertical drift profiles for seven limb/occultation records at ten NDACC lidar sites, six of which were also studied by Nair et al. (2012). The common vertical drift structure of the sounders noted at three stations (Andøya, Tsukuba and Dumont d'Urville) is indicative of features in the lidar time series, which may influence the network-averaged satellite drift analyses. Data from these three lidar sites are therefore rejected. The lidar record at Dumont d'Urville consists of two distinct periods (1991–1998 and 2008–now) with different data characteristics as a result of a complete system redesign (David et al., 2012). This may explain the large discrepancy between the drift results of historical and recent satellite record at this site (Fig. 4). In addition, we find that the drift results at Hohenpeißenberg for all recent sounders are significantly negative above about 25 km, while the results scatter around zero for two historic occultation instruments. Inspection of the time series indeed showed that the Hohenpeißenberg lidar reported more ozone for a few years after 2007. This station is hence discarded from the drift analyses of all satellite sounders operational during and after 2007 (Table 3). Similarly, the Table Mountain lidar (McDermid et al., 1990) measured higher ozone relative to satellite instruments during 2007–2008 (Nair et al., 2012). This bias disappeared in later years to leave the satellite drift estimates nearly unchanged. One exception is Aura MLS since the temporary lidar bias occurred close to the start of the mission. We nevertheless keep this lidar station in our analyses. A similar procedure was followed to discard about 20 ozonesonde records. For one satellite instrument we deviate from previous, standard selection of ground sites. SCIAMACHY drift results in the Arctic are very different from those in the rest of the atmosphere, especially for lidar. We believe this is a combined result of sampling and the seasonal cycle observed in the difference timeseries (Sect. 5). Therefore, all Arctic stations are excluded from the drift analysis of SCIAMACHY. Tables 1 and 2 list the stations used for the drift analysis (last column). Thanks to the pseudo-global coverage of the ozonesonde network, the network average should be a reasonably robust representation of the global satellite drift. Lidar network averages, on the other hand, are less representative of the global state and they are somewhat more sensitive to the station selection as well.

Figure S18 in Supplement shows how $\bar{\alpha}$ and $\sigma^*_{\bar{\alpha}}$ change when the discarded ground stations are included in the averaging procedure. Ozonesonde network-averaged drift and uncertainty change by less than $0.2\,\%\,\mathrm{decade}^{-1}$. The impact is a bit larger (< 0.5–$1\,\%\,\mathrm{decade}^{-1}$) for SCIAMACHY due to its peculiar data characteristics in the Arctic. Lidar network-averages are more sensitive to the selection of sites, especially for the recent satellite records. They differ by 1–$2\,\%\,\mathrm{decade}^{-1}$ above 25 km, mainly as a result of the inclusion of the Hohenpeißenberg data which systematically pulls the vertical drift profile towards more negative values. The impact of lidar site selection is much smaller for older records, less than $0.5\,\%\,\mathrm{decade}^{-1}$. Also the estimates of drift uncertainty are somewhat affected, but not as much as the actual drift values. Typically, the difference in uncertainty is less than $0.5\,\%\,\mathrm{decade}^{-1}$. Later on, we describe the remarkable agreement between the ozonesonde and lidar-derived drift results, strengthening the confidence in the stability of these ground networks (Fig. 5).

4.3 Results

Below we report on the vertical structure of the network-averaged drift estimates and their significance for each satellite record. We also mention (a) the smallest value of the 1σ regression uncertainty found across the network, (b) the typically found uncertainty, and, (c) the adjustment factor κ. This additional information illustrates the performance of the ground networks for this type of analysis. Main results are presented in Fig. 5 and summarised in Table 4.

4.3.1 SAGE II

The very long record of SAGE II, spanning 21 years, allows a detailed analysis of its stability. The smallest 1σ uncertainty derived at single sites in the ozonesonde and lidar networks is, respectively, $0.8\,\%\,\mathrm{decade}^{-1}$ and $1.6\,\%\,\mathrm{decade}^{-1}$. The average drift uncertainty over the ensemble of stations is $\sim 4\,\%\,\mathrm{decade}^{-1}$. The drift results are furthermore very consistent from one station to another, with a spread of 2–$3\,\%\,\mathrm{decade}^{-1}$ at 25 km (Fig. 4, grey line). The sonde and lidar derived estimates are statistically consistent as well. When aggregated over the entire ground network a significant SAGE II drift should be detectable at the 1–$2\,\%\,\mathrm{decade}^{-1}$ level, depending on altitude.

In the middle and upper stratosphere, between 20–40 km, the average drift is slightly negative except around ∼33 km (Fig. 5). The negative drift remains smaller than 1–2 % decade^{-1} and is not significant. At lower altitudes the drift becomes gradually more pronounced, but is never significant either as a result of the increased atmospheric variability or noise in the SAGE II record. We therefore conclude that the SAGE II record is stable relative to the ground measurements, at least within 2 % decade^{-1}.

4.3.2 SAGE III

SAGE III collected data for only 3.5 years, which excludes the upper stratosphere from our study as no lidar sites provide sufficient statistics. Between 20–30 km the minimal drift uncertainty is 6 % decade^{-1}, while that of most stations is easily twice as high.

SAGE III ozone decreases relative to ground measurements, by 2–6 % decade^{-1} in the middle stratosphere (MS) and more than 10 % decade^{-1} at lower altitudes (Fig. 5). The significance is by far insufficient however for a 2σ detection. The detection limit for the network-averaged drift is at best 6 % decade^{-1} between 20–30 km. In the lower stratosphere (LS) the threshold rapidly worsens to 10–30 % decade^{-1} due to the increased contribution of noise from natural variability and instrumental noise. We therefore conclude that SAGE III is stable within about ±10 % decade^{-1}, which is consistent an earlier report by Wang et al. (2006).

4.3.3 HALOE

The 14-year HALOE record allows a quite detailed study of the stability as well. The typical uncertainty at single stations is 5 % decade^{-1}, which is comparable to the variability of the spread between stations (Fig. 3, top left, light grey band). The 2σ detection threshold for the network average is 2–3 % decade^{-1} or more.

For altitudes above 100 hPa we observe a negative drift of about 1–7 % decade^{-1} (Fig. 5). The result is significant between 10–40 hPa for both the ozonesonde and the lidar comparisons. Figure 3 demonstrates that negative drifts are found across the entire ground network (left panel), all centered around the network-averaged value (right). At altitudes above 10 hPa and below 40 hPa the drift is less than ±5 % decade^{-1} with an uncertainty of 1.5–6 % decade^{-1} and hence not significant. No dependence on vertical coordinate or ozone unit was found for the HALOE drift results (Fig. 9), so these cannot be explained by drifting auxiliary data of the correlative records (Sect. 6).

Two earlier studies concluded that HALOE does not drift significantly relative to SAGE II, at least not more than ±10–15 % decade^{-1} (Morris et al., 2002; Nazaryan et al., 2005). Due to the longer data record considered here, the more frequent sampling and the stability of the ground networks we obtain a significant result already at the 2–3 % decade^{-1} level between 10–50 hPa. Our result is consistent with the earlier reports, although a direct comparison is not straightforward due to the different timespan and vertical coordinate (we come back to this in Sect. 6). From Fig. 5 we infer that the middle stratospheric drift of HALOE relative to SAGE II must range between 0 and −5 % decade^{-1}, which is comparable in sign and in magnitude with the −(0–10) % decade^{-1} reported by Morris et al. (2002, Fig. 4a) and −(2–4) % decade^{-1} by Nazaryan et al. (2005, Fig. 8).

4.3.4 UARS MLS

The UARS MLS record is somewhat short (less than 6 years) which limits the drift study especially at low altitudes and relative to the lidar instruments. Between 5–50 hPa (20–35 km) the single station drift uncertainty is 5 % decade^{-1} at best, but typically twice as large. When the results are averaged over the ground network the 2σ detection threshold is 4–8 % decade^{-1}. At other altitudes the threshold increases rapidly, by a factor of at least two.

For altitudes below 10 hPa the ozonesonde comparisons show a non-significant positive drift of 0–3 % decade^{-1} (Fig. 5). The drift relative to lidar, on the other hand, is negative but it is also not well constrained. As a result, the difference between the sonde and lidar-derived results is not significant. In fact, it is difficult to conclude anything from the lidar results; the results at different sites tend to be somewhat inconsistent, especially at altitudes above the 10 hPa level. While the upper stratospheric drift of UARS MLS goes up to +10 % decade^{-1} relative to the OHP and Table Mountain lidars, it goes down to −10 % decade^{-1} relative to the Mauna Loa and Lauder lidars. This necessitates a large χ^2-adjustment of $\kappa \simeq 2.5$ for lidar (Eq. (4) and Fig. S2 in Supplement) and results in a final uncertainty of about 10 % decade^{-1}. We conclude that between 10–50 hPa the UARS MLS instrument is stable within about ±5–10 % decade^{-1}, perhaps slightly worse. In the upper stratosphere the discrepancy between the lidar results is too large to assess the stability of UARS MLS.

We also note a dependence of the UARS MLS ozone drift results with profile representation due to an ascending drift in the accompanying GPH profile products (Fig. 9). More details and a recommendation to avoid such representation-dependences follow in Sect. 6.

4.3.5 Aura MLS

The stability of the Aura MLS instrument can be studied in great detail, thanks to its excellent temporal and spatial sampling. Single site drift uncertainty is at best 0.6 % decade^{-1} and 2 % decade^{-1} on average. Regression uncertainties are substantially smaller than the observed standard deviation of the drifts over the network, which is about 4–6 % decade^{-1} at altitudes above 50–100 hPa (Fig. 3, bottom). This leads to a considerable χ^2-adjustment (Fig. S2 in Supplement) of

$\kappa \simeq 2.5$ in the middle stratosphere (sonde) and $\kappa \simeq 3$ in the upper stratosphere (lidar). The resulting 2σ detection limit for network averages is 1–3 % decade^{-1} at altitudes below 5 hPa, and increases rapidly in the uppermost stratosphere.

In the upper and middle stratosphere the average drift is slightly positive, but generally not more than 1.5–2 % decade^{-1} (Fig. 5). Sonde and lidar derived results are very consistent. A significant negative drift seems to develop at altitudes below 100 hPa, which we think is due to an underestimation of the uncertainty. Indeed, obtaining realistic uncertainties at the level of a few % decade^{-1} level in the UTLS is a daunting task. We therefore conclude that Aura MLS v3.3 is stable in the entire stratosphere, certainly within 1.5 % decade^{-1} (MS) and 2 % decade^{-1} (US). Our ground-based estimates are consistent with earlier intercomparisons of Aura MLS, MIPAS (Eckert et al., 2014) and OSIRIS (Adams et al., 2014), indicating drifts between the instruments less than ±3–5 % decade^{-1}.

We will see later on that the above drift results differ from those in non-native profile representations, due to an overall descending drift of the Aura MLS GPH profiles (Fig. 9). This issue and a possible solution will be discussed in Sect. 6.

4.3.6 POAM II

The analysis of POAM II is extremely limited due to its infrequent sampling and short record, merely three years. The regression requirement of at least ten data points was met at just seven polar ozonesonde stations. There were not enough collocations with lidar instruments to study the upper stratosphere. Drift uncertainty is about 30 % decade^{-1} at most sites and 20 % decade^{-1} in the best case. The resulting 2σ detection threshold for the network average is 20–40 % decade^{-1} in the middle stratosphere. This is much larger than the observed drifts, which range from -15 % decade^{-1} at 20 km and 30 km to $+15$ % decade^{-1} at 25 km (Fig. 5). We conclude that the stability of POAM II is better than ±25 % decade^{-1} in the middle stratosphere.

4.3.7 POAM III

The POAM III data record spans 7.5 years and can therefore be studied in greater detail than that of its predecessor. In addition to the seven polar stations in the POAM II drift analysis, five ozonesonde sites at northern mid-latitudes provide a sufficiently sampled time series. Again, the regression was not feasible for lidar comparisons, limiting the altitude of our analysis to 30 km. The single station uncertainty is 4 % decade^{-1} at best and about 6 % decade^{-1} on average. When the results are averaged, the 2σ drift uncertainty becomes 4–8 % decade^{-1} in the middle stratosphere and rapidly grows to 10 % decade^{-1} at 15 km. Overall, POAM III seems to drift to lower ozone values between 20–30 km, at a rate of $-$(2–8) % decade^{-1} (Fig. 5). At lower altitudes the drift changes sign. None of our results are statistically significant. We conclude that POAM III is stable within, respectively, ±5 and 15 % decade^{-1} in the middle and lower stratosphere.

4.3.8 OSIRIS

The OSIRIS time series are densely sampled at many ground stations. In the middle and upper stratosphere the minimum drift uncertainty is 1.3 % decade^{-1} and typically amounts to 3–4 % decade^{-1}. The regression uncertainties do not fully explain the observed variability of 5–6 % decade^{-1} between stations above 20 km. The corresponding χ^2-adjustment factor κ is $\sim$ 1.5–2 for the sonde network and mostly less than 1.5 for the lidar network. The 2σ detection limit for the network average is 3 % decade^{-1} at 15 km, 1.6 % decade^{-1} between 20–30 km and 5 % decade^{-1} at 45 km.

In the lowermost stratosphere the OSIRIS drift relative to correlative measurements is negative, at most -5 % decade^{-1} and not significant (Fig. 5). There are clear indications of a positive drift between 15-35 km, of about 1–3 % decade^{-1}. While the sonde-derived result is significant (> 22 km), that is generally not the case for the lidar results (except between 28-34 km). In the upper stratosphere the positive drift becomes more pronounced and very significant above 37 km. Its presence is easily visible in the comparison time series, e.g. at the Observatoire de Haute-Provence lidar (Fig. 2). Around 42 km we find a > 2σ drift of +8 % decade^{-1} at three of the four best sampled lidar stations (Fig. 4, blue curves). Adams et al. (2014) reported a +(3–6) % decade^{-1} drift of OSIRIS relative to Aura MLS in the US, depending on how the Aura MLS data (pressure-VMR) are converted to the native OSIRIS system (altitude-number density). This is consistent with the 5 % decade^{-1} difference that we find between our lidar-based drift estimates for these two instruments.

In summary, OSIRIS ozone drifts very likely to higher values above 20 km. The drift is quite small up to 35 km and close to the 5 % significance threshold. In the upper stratosphere the presence of a +(5–8) % decade^{-1} drift is evident. The OSIRIS team has found that the drift in ozone is caused by a positive drift in the altitude registration. Efforts are under way to correct for this in the next data release.

4.3.9 SMR

Even though the SMR record spans 12 years and has good sampling properties, the ability to assess its stability is limited by the noise of the profiles. In Sect. 5 we show that the single SMR profile noise exceeds 20 % in the Tropics and 30 % at higher latitudes. This is substantially larger than for any other satellite record in this study. As a result, the drift uncertainty is at best 5–6 % decade^{-1} and typically $\sim$ 10 % decade^{-1} at individual ground sites. The regression uncertainties cover the observed drift variability across the ground network, so the χ^2-adjustment is close to one. In the

end, the 2σ threshold to detect averaged drifts ranges from 3–10 % decade^{-1} between 25–40 km.

The SMR profile drifts slightly to higher values in the middle stratosphere, although by no more than $+5$ % decade^{-1} which is insignificant (Fig. 5). Above 30 km the drift changes sign and increases rapidly in magnitude, reaching -12 % decade^{-1} around 40 km. Due to the large single-profile noise the negative drift is only significant at 2σ level between 40–43 km. Our result contrasts with satellite intercomparisons by Jones et al. (2009) which indicated an insignificant positive drift of SMR relative to a multi-satellite average in the upper stratosphere. This difference may be due to a shorter period (2001–2007) or the different data versions, and deserves further study. In the meantime, we conclude that SMR is stable within $\pm$6–8 % decade^{-1} over most of the stratosphere. SMR ozone trends in the uppermost stratosphere, however, should be interpreted cautiously as they possibly underestimate the actual trend by a fair amount.

4.3.10 GOMOS

The constraints on the stability of GOMOS are weaker than for its contemporary limb sounders, due to sparser sampling and, below $\sim$20 km, the larger noise of GOMOS. The latter is, for instance, clear from the comparison time series at the Payerne ozonesonde station (Fig. 2). In the middle stratosphere, the drift variability between stations is about 10 % decade^{-1}, which is larger than the uncertainties at individual sites, about 3 % decade^{-1} at best and 7 % decade^{-1} in general. The χ^2-adjustment increases the uncertainty of the network averages by $\kappa \simeq 1.5$. The resulting 2σ detection threshold is 3–5 % decade^{-1} between 20–40 km and raises rapidly in the lower stratosphere, e.g. to $\sim$ 12 % decade^{-1} at 15 km.

Below 25–30 km a pronounced negative drift develops with decreasing altitude, from -1 % decade^{-1} at 30 km to -4 % decade^{-1} at 20 km (Fig. 5). The results for the ozonesonde and lidar networks are qualitatively and quantitatively consistent, the latter being less significant below 22 km. GOMOS drift estimates are close to the 2σ threshold between 15–25 km. At lower altitudes, the significance decreases due to marked increased noise. Our observations are qualitatively consistent with Tegtmeier et al. (2013, Fig. 8) or Nair et al. (2011). The latter reported a drift of up to -18 ± 8 % decade^{-1} (1σ uncertainty) near 20 km relative to the lidar at Observatoire de Haute-Provence (43.9° N, 5.7° E). In the upper stratosphere our lidar results are scattered, but they point on average to a positive drift above 35 km (Fig. 4, green line). The maximum drift is $+3$ % decade^{-1} at 45 km, which remains well below the 2σ threshold.

4.3.11 MIPAS

We only consider profiles from 2005–2012 (optimised resolution period, OR) in the nominal observation mode, since other MIPAS data is less recommended for use in long-term studies. Nevertheless, the stability can still be studied down to several % decade^{-1} thanks to the good sampling properties of the instrument. In the middle and upper stratosphere, the smallest single-site regression uncertainty is 1.5 % decade^{-1} and typically $\sim$3 % decade^{-1}. These errors do not fully cover the observed variability between sonde stations. They are therefore scaled by a factor of $\kappa \simeq 2$ between 3–50 hPa and $\simeq$1 at altitudes below 100 hPa. The resulting 2σ detection limit for the network average is 2–4 % decade^{-1} between 10–100 hPa and 5–9 % decade^{-1} in the upper stratosphere.

No significant drift is observed in the MIPAS OR profiles, they are stable relative to the ground-based networks over the entire considered altitude range. Drift estimates are less than ±2 % decade^{-1} in the middle and upper stratosphere, and less than ±4 % decade^{-1} at lower altitudes (Fig. 5). Eckert et al. (2014), on the other hand, noted clear negative drifts in the upper stratosphere between MIPAS data retrieved by the Level-2 processor at Karlsruhe Institute of Technology and Aura MLS (0.2–0.3 ppmv decade^{-1}, or $\sim$3–5 % decade^{-1}) or OSIRIS (0.3–0.6 ppmv decade^{-1}, or $\sim$5–10 % decade^{-1}). The seemingly contrasting results from both analyses are nevertheless in good agreement. We deduce from the lidar-based drift estimates that the relative drift between MIPAS and Aura MLS or OSIRIS would be, respectively, $-$(2–5) and $-$(3–10) % decade^{-1} for altitudes above 5 hPa (Fig. 5).

Our drift results are generally not applicable for trend analyses which include MIPAS data prior to 2005 (full resolution period, FR). The FR data are biased relative to the OR profiles (Ceccherini et al., 2013), which will introduce an (altitude-dependent) systematic uncertainty in trend analyses if not accounted for. Eckert et al. (2014) overcome this issue by including the FR-OR bias as a free parameter in the regression model.

4.3.12 SCIAMACHY

The excellent sampling of SCIAMACHY allows us to probe its stability down to 0.8 % decade^{-1} at some ground sites, and on average down to $\sim$2 % decade^{-1}. Again, these statistical uncertainties do not cover the variability of 6 % decade^{-1} observed between the stations, leading to a $\kappa \simeq 2 - -2.5$ adjustment over most of the middle stratosphere. The drift averages become significant when they cross the 2–6 % decade^{-1} bar in the middle and upper stratosphere.

Below 30 km the SCIAMACHY profiles drift to higher values relative to ozonesondes and lidars (Fig. 5). The drift is nearly independent of altitude and amounts to about

$+2\,\%\,\mathrm{decade}^{-1}$. The sonde results surpass the 2σ threshold, but those derived from lidar observations do not. The drift has opposite sign above 30 km and becomes rapidly highly significant at all lidar sites (Fig. 4, purple line). It reaches maximal significance, more than 5σ, around 38 km with a magnitude of $-9\,\%\,\mathrm{decade}^{-1}$. Figure 2 illustrates the negative drift at 38 km in the comparison time series for the Mauna Loa lidar (19° N, 156° W).

These results clearly show that SCIAMACHY trend results should be interpreted very cautiously in the upper stratosphere, and likely at lower altitudes as well. For instance, the large negative drift in SCIAMACHY US ozone explains, at least partially, the more negative trends derived from the IUP Bremen v2.5 data set than those found for Aura MLS and OSIRIS (Gebhardt et al., 2014). While latter authors consider a different SCIAMACHY Level-2 processor than here, there have been reports of a negative drift of $5\,\%\,\mathrm{decade}^{-1}$ at 30–40 km for the IUP Bremen processor as well (Tegtmeier et al., 2013; Lambert et al., 2014). Both SCIAMACHY processors digest the same Level-1 data, so it is plausible that the degradation of the instrument is not entirely corrected for by the calibration scheme (Krijger et al., 2014).

4.3.13 ACE-FTS

The solar occultation instruments onboard SCISAT sample mainly high latitudes. We limit our stability study of ACE-FTS to the lower and middle stratosphere, since there is only one lidar site with a sufficient number of co-locations. The best single-site drift uncertainty is $3\,\%\,\mathrm{decade}^{-1}$, whereas it amounts to about $10\,\%\,\mathrm{decade}^{-1}$ in general, close to the observed variability between stations. The observed drift is mostly negative, less than $5\,\%\,\mathrm{decade}^{-1}$, which is consistent with the no-drift hypothesis (Fig. 5). The ACE-FTS data record can be considered stable to within about $5\,\%\,\mathrm{decade}^{-1}$. A more precise analysis will be possible once the ACE-FTS profiles taken after September 2010 are included in the analysis.

4.3.14 MAESTRO

The uncertainty on the stability of the MAESTRO record is slightly poorer than that of ACE-FTS. The larger single-station uncertainties, at least $5\,\%\,\mathrm{decade}^{-1}$ and typically $12\text{–}14\,\%\,\mathrm{decade}^{-1}$, lead to a 2σ detection threshold at $6\text{–}8\,\%\,\mathrm{decade}^{-1}$ and $6\text{–}25\,\%\,\mathrm{decade}^{-1}$ in the middle and lower stratosphere, respectively. The results never cross these thresholds: below 20 km we find a drift between -7 and $+10\,\%\,\mathrm{decade}^{-1}$, above 20 km the drift is mainly positive and about $2\text{–}3\,\%\,\mathrm{decade}^{-1}$ (Fig. 5). Hence, the MAESTRO record is considered stable within $\pm 6\text{–}10\,\%\,\mathrm{decade}^{-1}$. Again, as for ACE-FTS, the uncertainty will decrease once the post-September 2010 profiles will be added to the analysis.

5 Bias and short-term variability

After studying decadal stability, we address the overall bias and short-term variability and search for patterns in altitude, latitude and season. As in the previous section, we focus here on the individual satellite records in their native profile representation. Later on we expand the discussion to the consistency between profile representations (Sect. 6) and between satellite records (Sect. 7).

5.1 Methodology

Again, robust statistics are adopted that protect against outliers. We define the bias $b(l)$ as the median of the difference distribution at grid level l

$$b(l) = \mathrm{Q}_{50}(\Delta x_i(l)), \tag{5}$$

where i runs over the pairs in the comparison sample. The 68 % interpercentile of the difference distribution

$$s(l) = \frac{1}{2}\Big[\mathrm{Q}_{84}(\Delta x_i(l)) - \mathrm{Q}_{16}(\Delta x_i(l))\Big] \tag{6}$$

is referred to as comparison spread s. We stress that s should not be confused with an estimate of the precision of the satellite data, as other, non-negligible terms enter the comparison error budget. These include the precision of the ground-based data and random uncertainties in the metrology of the comparison related to the difference in sampled air masses (Sect. 3), but also any long-term time dependence of the bias. In principle a similar remark is also valid for the bias b, but systematic uncertainties in the metrology of the comparison are expected to play a smaller role, except perhaps in the UTLS due to low ozone abundances and above 30 km due to the different sampling by lidar and a few satellite instruments of the diurnal cycle (Sect. 3).

5.2 Results

The vertical and meridional structure of bias and comparison spread relative to ozonesonde measurements is shown in Figs. 6 and 7. Since there is more resemblance between the instruments, we only show a few typical cases for the comparison spread. Table 5 summarises the bias estimates in four layers of the atmosphere. In the Supplement we provide vertical profiles of bias and spread from comparisons to ozonesonde and lidar observations in five latitude bands (Figs. S3–S16). In addition, for selected instruments, there are supplementary figures for the dependence of data quality on solar occultation type (Fig. S19) and month (Fig. S20).

5.2.1 SAGE II

Between 20–40 km SAGE II ozone remains mostly within $\pm 3\,\%$ of the correlative measurements. Above 30–35 km, however, sunrise profiles have a $\sim 4\,\%$ more negative bias

relative to lidar than sunset profiles (Fig. S19 in Supplement). This confirms, qualitatively, earlier reports of 8–10 % smaller sunrise concentrations than at sunset in the middle and upper stratosphere (Kyrölä et al., 2013; Damadeo et al., 2014; Sakazaki et al., 2015). In the lowermost stratosphere, and below, ozone is underestimated by up to 10–15 %. The spread in the comparisons is lowest between 25 and 40 km and shows poleward increases, 5 % at the Equator and ∼ 10 % at the high latitudes. Below 20 km the observed spread increases rapidly to 20–30 %, and especially under Antarctic ozone hole conditions.

5.2.2 SAGE III

The stratospheric bias of SAGE III is mostly less than ±3 %, comparable to that of its predecessor. Ozone is generally slightly overestimated except in the Arctic between 10–35 km and below ∼ 15 km at mid latitudes. The latter contrasts with a high bias up to 10 % seen at 13 km by Wang et al. (2006) for an earlier version of the data set. It is not clear whether the SAGE III sunrise and sunset profiles are biased relative to each other. Figure S19 (in Supplement) shows that the bias relative to lidar is ∼5 % more positive for sunrise measurements above 30 km. However, it is not possible to attribute this to diurnal variation since the type of occultation depends on the hemisphere (sunset in North, sunrise in South) and there may be a meridian structure in the instrument bias field (Fig. S4). The short-term variability seems a few percent better than that of SAGE II, i.e. about 5 % at mid-latitudes and 8 % in the Arctic. Below 20 km and above 35–40 km the variability in the comparisons increases markedly.

5.2.3 HALOE

In the upper stratosphere and tropical middle stratosphere HALOE overestimates ozone by up to 3 %. In contrast, a negative bias is noted over the rest of the atmosphere. In the middle stratosphere it is not more than 5 % but it decreases rapidly at altitudes below 50 hPa, reaching at least 25 % at 200 hPa. The variability in the comparisons is similar to that from the SAGE instruments, ranging from 5–10 % in the middle and upper stratosphere, and peaking at 30–40 % around the tropopause. During the Antarctic ozone hole season, the volume mixing ratios are overestimated by 25 % and the spread increases to 35 %. Our results are consistent with earlier satellite and ground-based studies (Morris et al., 2002; Nazaryan et al., 2005). Sakazaki et al. (2015) reported a 2–5 % positive bias of sunset relative to sunrise occultations above 40 km. The lidar-based analysis seem to confirm this, differences between both occultation types are less than 2 % below 40 km and somewhat higher in the uppermost stratosphere (Fig. S19).

5.2.4 UARS MLS

Our findings corroborate most of those by Livesey et al. (2003): (a) at altitudes above 50 hPa UARS MLS slightly overestimates ozone by up to 5 %, (b) the 68 hPa and 100 hPa levels exhibit larger biases up to 10 %, and (c) the bias peaks at 68 hPa. However, negative biases up to 5 % are seen between 10–50 hPa relative to southern ozonesondes and at altitudes above 5 hPa relative to northern lidars. The short-term variability is similar to the previous records, but reaches the 5–10 % range somewhat higher up in the middle stratosphere, around 20 hPa. At lower altitudes the comparison spread increases fast, maximizing at more than 40–50 % at the tropopause. We noted furthermore that the UARS MLS bias depends on the profile representation if one uses the GPH and temperature data included in the MLS product to perform conversions. This will be discussed in more detail in Sect. 6.

5.2.5 Aura MLS

Aura MLS ozone remains within ±3 % of correlative measurements between 5–50 hPa, except in the Arctic where a negative bias of 5 % is noted. The most striking bias characteristics are the stationary vertical oscillations found in the finer vertical retrieval grid results (version 3.3/3.4 data, see Livesey et al. (2013a)). They are very pronounced in the tropical UTLS where the amplitude reaches 10–15 %, but also extend to higher latitudes and altitudes, with amplitudes of 3–5 %. The previous data release, v2.2, has a coarser grid in the UTLS and displays fewer oscillations. The recent new release of Aura MLS data (version 4.2) mitigates these oscillations to some extent (Livesey et al., 2015). The comparison spread shows that the single-profile precision is better than 4–7 % in the middle and upper stratosphere, and starts to degrade for altitudes below 50 hPa. The Aura MLS bias furthermore depends on the profile representation if one uses the GPH and temperature data included in the MLS product to perform conversions. We come back to this in Sect. 6.

5.2.6 POAM II

We observe a negative bias of about 5–10 % between 20–30 km, which becomes rapidly more pronounced at lower altitudes in the Antarctic. This is consistent with earlier satellite and ground-based studies (Rusch et al., 1997; Deniel et al., 1997; Danilin et al., 2002). In the northern lower stratosphere, however, the negative remains less than 5 %. Above 30 km, there is a positive bias of 5–10 % or more relative to the polar lidars. The small lidar comparison sample did not allow us to study sunrise versus sunset results. As for SAGE III, the observed differences (Fig. S19) could also be due to a meridian dependence of the instrument bias since the occultation type changes with hemisphere. The comparison

spread is 5–10 % in the middle and upper stratosphere, and increases below 20 km.

5.2.7 POAM III

The POAM III bias is less than 5 % in the middle stratosphere and upper stratosphere, and has a negative sign between 18–30 km and positive elsewhere. Here, the spread in the comparisons is also similar to its predecessor, ranging between 5 and 10 %. In the lower stratosphere there is an overestimation of at least 10 %, and, again, the spread is more pronounced. Our results corroborate the findings of Randall et al. (2003). Unfortunately, the small comparison sample does not allow us to verify their report of a negative bias of up to 5 % of MS and US sunrise data (taken in the Northern Hemisphere) relative to sunset profiles (in SH).

5.2.8 OSIRIS

Our ground-based bias results are very consistent with those of satellite intercomparisons (Adams et al., 2013, 2014). OSIRIS ozone remains mostly within ±4 % of correlative measurements above 20 km, but two features stand out. First and foremost, a marked peak in bias around 22 km is seen at all latitudes which is possibly related to biases in the aerosol retrieval preceding the ozone retrieval (Adams et al., 2014). The comparison to lidars in the Tropics and Southern Hemisphere shows a second jump towards a persistent 5 % positive bias, occurring between 30–35 km. Such a feature is not seen in the Northern Hemisphere. In the lower stratosphere, below 20 km, OSIRIS underestimates ozone by 5–10 % at mid and high latitudes and by more than 15 % in the Tropics. Comparison spreads range from 6 to 11 % between 20 and 35–40 km. In the UTLS these increase to 20–40 % at 15 km, depending on latitude.

5.2.9 SMR

Our analysis confirms earlier reports (by e.g. Urban et al., 2005; Jones et al., 2007; Jégou et al., 2008) of a systematic underestimation by 5–10 % in the upper and (most of the) middle stratosphere. The bias changes sign at lower altitudes and peaks at +5 to +10 % around 20 km. The most notable characteristic is the high comparison spread. It increases in the middle stratosphere from 20 % to 30 % between the Tropics and the polar regions, and becomes even larger at other altitudes (Fig. 7). The poor single-profile precision is caused by the low signal-to-noise ratio for the 501.8 GHz line used for the ozone retrievals. Better precision can be obtained by averaging the profiles in the logaritmic VMR domain (Urban et al., 2005). Alternatively, one could use the SMR ozone products from the stronger 544.6 GHz band, these are clearly less noisy though exhibit larger biases (Hassler et al., 2014).

5.2.10 GOMOS

The GOMOS ozone bias above 20 km is generally less than ±3 %. The exception is the Arctic where a −7 % bias is found relative to ozonesonde and lidar at 25 km and again at 40 km. This is in agreement with earlier analyses by van Gijsel et al. (2010). Another notable feature is that the sign of the bias in the extratropical UTLS is opposite in both hemispheres. It reaches −20 % in the North and +20 % in the South at 10 km. The larger biases below 20 km are due to the interference of ozone and aerosol retrievals with aerosol models (Tamminen et al., 2010). In the middle and upper stratosphere the comparison spread ranges from 6 % in the Tropics to 11 % at high latitudes (Fig. 7). Below 20–25 km GOMOS data becomes notably more noisy; at 15 km the observed spreads amount to 25–50 %, as a result of the increasing opacity of the atmosphere. Theoretically it is expected that profile quality depends on star properties such as magnitude and temperature. However, our analysis confirms (not shown here) an earlier claim by van Gijsel et al. (2010) that this is not the case when the recommended screening procedure is applied. The illumination condition of the occultation is clearly a more determining factor, with dark limb profiles offering best data quality.

5.2.11 MIPAS

Due to changes in instrument and retrieval set-up there is an altitude-dependent bias between the first (2002–2004) and later years of the mission of up to 5 % (Ceccherini et al., 2013; Eckert et al., 2014). Our analysis covers the 2005–2012 period only and corroborates earlier findings for the operational and several alternative MIPAS Level-2 processors (e.g. Cortesi et al., 2007; Laeng et al., 2014, 2015). MIPAS OR profiles overestimates ozone systematically over most of the stratosphere, except in the Arctic. At mid and low latitudes there are two bias peaks of +(5–10) % around 50 hPa and 5 hPa. At other pressure levels the bias remains below 5 %. At the bottom of the profile, for $p > 200$ hPa, ozone is overestimated by at least 20 %. In the Tropics, the bias briefly flips sign between 50–200 hPa, where a very negative bias is found. Between 2–50 hPa, the observed spread ranges from 4 % in the Tropics to 8 % at higher latitudes. Again, in the UTLS a sharp increase is observed (Fig. 7). We also noted a dependence of the MIPAS bias on the profile representation when the pressure and temperature data included in the operational product are used to perform conversions. More details follow in Sect. 6.

5.2.12 SCIAMACHY

The SCIAMACHY bias is clearly positive over most of the atmosphere and manifests an intricate structure in altitude, latitude and season. The agreement with ozonesonde and lidar is better than 10 %, and best at northern mid-latitudes

(between 0 and +5 % over 15–40 km). However, the bias easily reaches +10–15 % over a large part of the stratosphere, stretching from 30° N–60° S (> 25 km) to 60° S–90° S (> 30 km). Similar results were obtained by Tegtmeier et al. (2013) for an alternative Level-2 processor developed by IUP Bremen. Arctic profile data quality is particularly peculiar (Fig. S20 in Supplement). There is a clear vertical dependence of the bias, peaking at +10 % around 20 km and −10 % at 15 and 30 km. Also, both bias and comparison spread vary strongly with season. The bias at 20 km reaches a maximum of +25 % during boreal winter and a minimum of 0 % in summer. Similarly, the mean comparison spread is about 20 %, but it peaks at 30 % in winter and shrinks to 10 % in summer. At other latitudes the observed spread is never below ∼ 10 %. Furthermore, in Sect. 6 we will show that the SCIAMACHY bias depends on the profile representation if the pressure and temperature data included in the SCIAMACHY product are used to perform conversions.

5.2.13 ACE-FTS

ACE-FTS ozone remains generally within about ±3 % from ground-based measurements over the entire stratosphere. The bias is negative relative to Arctic ozonesondes, everywhere else ozone mixing ratios are overestimated. Above 30 km, the comparisons to mid-northern and high-southern lidars indicate a slightly larger positive bias, but not more than 5 %. The relative bias only exceeds 5 % a few km above the tropopause. These observations are in line with other studies (Dupuy et al., 2009; Waymark et al., 2013). Sakazaki et al. (2015) recently reported sunrise-sunset biases in the upper stratosphere of 2–5 % above 40 km. Figure S17 shows differences of similar magnitude between the lidar bias results for both occultation types, but with opposite sign and penetrating deep into the middle and lower stratosphere. These results are clearly due to statistical fluctuations in the small co-location sample. The ACE-FTS record performs also well in terms of short-term variability. The comparison spread is at most 7 % (10 %) at high latitudes in the middle (upper) stratosphere. It increases strongly below 20 km.

5.2.14 MAESTRO

The MAESTRO profiles exhibit typically a negative bias in the Northern Hemisphere (3 to 6 %) and a positive bias in the Southern Hemisphere (0 to 10 %). Ozone is clearly underestimated below 15 km, by at least 20 % at all latitudes. These findings confirm those by other authors (Kar et al., 2007; Dupuy et al., 2009). Earlier reports of a negative bias of up to 20 % between sunrise and sunset measurements above 35 km can not be confirmed or excluded. Our lidar-based bias results for sunset and sunrise data differ less than 5 % in the upper stratosphere (Fig. S19). Nonetheless, the comparison sample is quite small so it does not necessarily provide a representative picture. The observed comparison spreads range from 7 % at mid-latitudes to 10 % in the Arctic.

6 Impact of auxiliary data on non-native representations

So far, we have considered the quality of the satellite records in their native profile representation. But a user may actually desire another representation depending on his/her application (e.g. model comparisons, merging or assimilation of different records). In this case coincident altitude, pressure and/or temperature profile data are necessary for the conversion between ozone VMR and number density or between altitude and pressure. Users may prefer measurements, climatologies or reanalysis fields, all of which bring along uncertainties (e.g. Thorne et al., 2011; Seidel et al., 2011; Stauffer et al., 2014; Simmons et al., 2014). These ultimately add uncertainty $\mathbf{S}_{\mathrm{auxiliary}}$ to the transformed ozone profile, which may have structure in space (altitude, latitude) and time (short- and long-term). Moreover, the currently observed negative trend of 1 K decade^{-1} in upper stratospheric temperature data already leads to representation-dependent differences of up to 1 % decade^{-1} in the ozone trends (McLinden and Fioletov, 2011). It is important to realise that a drift in temperature data will, in a similar fashion, introduce extra (altitude-dependent) drift in non-native ozone representations. Here, we consider the auxiliary data provided in the satellite data files, see Table 3. The auxiliary profiles for ground-based data are taken either from actual measurements (ozonesonde: interfaced radiosonde), or from reanalysis fields (lidar: ERA-Interim).

At altitudes below about 35 km (∼ 5 hPa) there is generally no clear change in bias or comparison spread (both < 1 %) and drift (< 1 % decade^{-1}) after the conversion to another representation. There is, therefore, no considerable difference in bias, short-term variability or long-term stability of the auxiliary data for most satellite and ground-based profiles. Examples are shown for SAGE II bias (Fig. 8) and HALOE and OSIRIS drift (Fig. 9). Complete information for all sounders can be found in the Supplement (Figs. S3–S16). Observations of upper stratospheric temperature are generally less consistent (Simmons et al., 2014), so it is not surprising to find considerable changes in ozone bias (up to ∼ 5 %) or drift (up to ∼ 5 % decade^{-1}) around 45 km (∼ 1 hPa).

For a few records we find clear indications that the accompanying auxiliary data have a more important impact than the numbers stated before. MIPAS bias changes by about 3 % when switching between VMR and number density, except between 25–30 km (∼ 10–20 hPa), see Fig. 8 and Fig. S13 in Supplement. The effect is slightly more pronounced in the Tropics and slightly less in the polar regions. Interestingly, transforming the vertical coordinate does not influence the ozone bias even though there is a ∼ 200 m negative bias in MIPAS altitude. This indicates that the averaging ker-

nel smooths out the effect of altitude offsets. The observed dependence on ozone unit is not caused by the conversion procedure of the averaging kernel, since a similar depdendence is seen in MIPAS comparisons to non-smoothed cor-relative data (not shown here). The SCIAMACHY bias depends on both vertical coordinate and ozone unit over the entire stratosphere, by 3–5 % (Fig. 8 and Fig. S14), likely as a result of uncertainties in the McLinden p/T climatology. Both MLS records on the other hand exhibit clear represen-tation dependences of the drift (Fig. 9 and Figs. S6–S7), and to a lesser extent also of the bias (although the ozonesonde and lidar results are somewhat discrepant, Fig. 8). The Aura MLS drift changes by about 3 % decade^{-1}, similar to earlier reports (Adams et al., 2014). The dependence has opposite sign for UARS MLS and is more pronounced, up to $\sim 10\,\%\,\mathrm{decade}^{-1}$. These observations are consistent with the known drifts in absolute pointing of the MLS records. Whereas UARS MLS geopotential height profiles drift upwards, by $\sim 1000\,\mathrm{m\,decade}^{-1}$ (Livesey et al., 2003), the Aura MLS v3.3/v2.2 GPH data drift downwards by $\sim 120\,\mathrm{m\,decade}^{-1}$, especially between 2005–2009 (Livesey et al., 2015). Obviously, if a more stable and less biased source of auxiliary data were used for the conversion, the reported issues for MIPAS, SCIAMACHY and the MLS records could be easily avoided. Our results suggest that radiosonde data, reanalysis fields by ERA-Interim (lidar) and MERRA (SAGE II), and ECMWF operational data (SMR, OSIRIS, GOMOS) allow for consistent conversions.

7 Consistency between satellite records

Until now we discussed each satellite Level-2 data set individually. Here, we take advantage of the specific design of the analysis to compare the satellite records directly. What follows is an evaluation of their mutual consistency in terms of bias (Fig. 10), short-term variability (Fig. 11), decadal stability (Fig. 12) and auxiliary data. In the context of the SI2N initiative an extensive literature review was performed of the ground-based validation and satellite intercomparison studies (Lambert et al., 2016). We refer the reader to this work for a more in-depth discussion of the global picture that emerges from the different studies.

7.1 Bias

Figure 10 shows a superimposed view of the vertical structure of satellite bias in five latitude bands, in the native representation of each satellite record. The smallest biases and best mutual consistency are found between 20–40 km ($\sim$2–50 hPa). Here, satellite and ground-based measurements mostly agree within 5 % or better (grey shaded area). Furthermore, the inter-satellite bias is not more than about 5 %. This illustrates the excellent consistency of all satellite and ground-based records in this part of the atmosphere. The consistency appears slightly poorer in the uppermost stratosphere (above 40 km/$\sim$2 hPa), perhaps due to the lower ozone abundances or due to larger systematic uncertainties in the lidar measurements. In the lower stratosphere and below the tropopause there is a clear degradation of the percentage bias and consistency, due to declining ozone levels and increasing interference by clouds and aerosols (Wang et al., 1999, 2002; Randall et al., 2003). The bias relative to sondes easily reaches 15 % and more, and the inter-satellite biases can be more than twice as large. Exceptions to this general picture are POAM II (dashed green) and especially SCIAMACHY (solid yellow). POAM II ozone is systematically low by about 5–10 % in the middle stratosphere, except in the Arctic. The SCIAMACHY bias reaches 10 % and more over a large part of the stratosphere, with a peculiar meridional structure, and a seasonal dependence that is very pronounced in the Arctic (Fig. S20 in Supplement). Section 5 presented noteworthy bias features also for other records, but of smaller magnitude and at smaller atmospheric scales: SMR (crosses the $-5\,\%$ threshold above 35 km or $\sim$5 hPa), Aura MLS (distinctive vertical oscillations in the UTLS), MIPAS (persistent positive bias of $\sim 5\,\%$), OSIRIS (sudden anomaly in bias around 22 km) and GOMOS (larger negative bias in Arctic).

Some bias features in Fig. 10 are common to the satellite measurements, and, hence, possibly relate to the ground-based data quality. Perhaps the most striking, and not understood at the moment, is that the Arctic middle stratospheric bias is negative for most satellite records, relative to both sonde (8 stations) and lidar (3 sites). This may indicate that the Arctic ground-based ozone values are too high. Secondly, there is a systematic positive upper stratospheric bias at tropical and southern mid-latitudes, possibly caused by a small negative bias of the dominating lidar record (Mauna Loa and Lauder). These $\sim 3\,\%$ biases remain within the systematic uncertainty due to uncertainties on the absorption cross-sections used for the lidar retrievals. Thirdly, a $\sim$10–15 % negative bias seems present in the early Dumont d'Urville lidar record (1991–1998) (Godin et al., 2001). Indeed, all satellite records that started before 1999 (dashed lines) are biased high with similar magnitude in the Antarctic middle and upper stratosphere, while that is not the case for the more recent comparisons at this lidar site (2008–2013, solid). And finally, we systematically note a curved vertical structure of the bias relative to ozonesondes: sonde ozone values are decreasing by up to 5 % between 25 km and the top of the profile. This may be related to an incomplete sonde correction scheme for the decrease in pump performance or for the increase in vertical registration error due to biases in the pressure readings (Stauffer et al., 2014). Apart from these differences, the ozonesonde and lidar results are highly consistent, highlighting the suitability of these ground-based networks as a transfer standard.

7.2 Comparison spread

The comparison spread results in Fig. 11 are more straightforward than those of the bias. There is a consistent dependence on latitude and altitude for all records. Between 20–35/40 km (∼ 2/5–50 hPa) the spread ranges between 5–12 % (grey shaded area) and increases slightly from the Tropics towards the poles, qualitatively consistent with a larger co-location mismatch uncertainty due to higher natural variability at high latitudes. Above 35–40 km (∼ 2–5 hPa) ozone levels decrease and the precision of the lidar measurements degrades, leading to a ∼ 5 % and more increase in comparison spread. Below 20 km (∼ 50 hPa) the spread increases rapidly, easily more than 40 % at the tropopause, due to the higher natural variability. But here the lower signal to noise ratio (clouds, aerosols) plays a role as well and differences in comparison spread between records become obvious. GOMOS and UARS MLS appear less sensitive to ozone in the lower stratosphere. The most precise measurements over the entire stratosphere, on the other hand, are made by ACE-FTS, Aura MLS and MIPAS, although the comparison spread results for the latter two records may include a smaller co-location mismatch component due to the tighter time window (6 h instead of 12 h). SCIAMACHY and SMR are clearly different. The single-profile variability in the SMR comparisons is more elevated over the entire stratosphere (20–30 %). For SCIAMACHY this is seen (Fig. S14 in Supplement) in the upper stratosphere (10–15 %), and particularly in the Arctic (25–40 %) where a clear anomaly is discerned around 25 km, together with a very strong seasonal dependence (10 % in boreal summer, more than 30 % in winter). During the Antarctic ozone hole season, the extremely low ozone conditions inflate the comparison spread of all records to 40 % or more around 20 km (not shown here in detail). The low signal to noise ratios thus pose a real challenge for all limb/occultation sounders.

7.3 Decadal stability

Figure 12 presents a superimposed view of the vertical structure of the ground-network averaged decadal stability of all satellite records[2], in their native representation. The drift relative to ground observations is generally not significant and less than 5 % decade^{-1} in the middle and upper stratosphere, for some records even better than 3 % decade^{-1} over a large part of the stratosphere. The relative drift between satellite records can be twice as large however. A few records deviate from this general tendency. Either seemingly so because of large drift uncertainty (UARS MLS, SAGE III, POAM II), or because of the presence of a significant drift (HALOE between 20–30 km, SCIAMACHY between 32–42 km, OSIRIS between 36–44 km). The GOMOS (below 25 km) and SMR (above 35 km) records may also drift, although the results are close to the detection threshold for these instruments. Another peculiarity is the possible presence of a common, weak vertical dependence of the drifts in the middle stratosphere. These tend to become gradually more positive with increasing altitude, by 1–2 % decade^{-1} between 20–30 km, see Fig. 5 (e.g. SAGE II, Aura MLS, OSIRIS, GOMOS). This unexplained feature is observed independently of satellite record, or of type of correlative instrument, and deserves further study.

7.4 Impact of auxiliary data

Satellite ozone profile data quality is generally not affected by the conversion to another representation with the help of the accompanying pressure and temperature profiles. Bias, spread or decadal stability typically change, respectively, by less than 1 %, 1 % or 1 % decade^{-1} in the lower and middle stratosphere, and somewhat more in the upper stratosphere. This demonstrates the good mutual consistency of the meteorological data by ozonesonde, MERRA and ERA-Interim. The exceptions are MIPAS and SCIAMACHY (3–5 % change in bias) and the Aura MLS and UARS MLS records (respectively 3 and 10 % decade^{-1} change in drift). Obviously, the introduction of these artificial effects can be avoided by using less biased or more stable sources of auxiliary data.

8 Discussion

The patterns in bias, short-term variability and decadal stability of the Level-2 ozone profile records identified in the preceding sections will affect higher-level products if not properly accounted for. Many studies within the community are based on gridded Level-3 data (e.g. monthly zonal means from single or a combination of instruments) or assimilated Level-4 fields. In this section we discuss the relevance of our Level-2 assessment for the construction and analysis of such derived records, and focus in particular on implications for recent ozone profile trend assessments.

8.1 Can end-user requirements be verified?

We start the discussion by reflecting on the requirements of end users. Naturally, these depend on the envisaged application, so various sets of requirements have been drafted by the community[3]. We focus here on climate applications which rely on stable data sets spanning multiple decades on a global scale. The Global Climate Observing System (GCOS), for instance, requests an accuracy (Joint Committee for Guides in Metrology, 2012) better than 10 % in the UTLS and 5–20 % above, and a stability better than 1 % decade^{-1} (GCOS, 2011). Within ESA's Climate Change

[2]To avoid clutter in Fig. 12 SAGE III and POAM II are not shown in the panels on the left. These can be seen in Fig. 5.

[3]For an overview, see http://www.wmo-sat.info/oscar/variables/view/108.

Initiative program (Ozone_cci) similar requirements were set for accuracy (< 8–15 %) and somewhat looser targets for stability (< 1–3 % decade^{-1}) (van der A et al., 2011).

In practice, the accuracy and stability of a particular record can of course only be tested to a level determined by the accuracy and stability of the reference data and by constraints from the metrology of the comparison. From Figs. 10 and 11 we conclude that ground-based studies are indeed able to verify an accuracy of 5–10 %, and resolve altitude-latitude-season patterns, in the middle and upper stratosphere. This is much more challenging in the UTLS, where uncertainties in the metrology of the comparison become important due to increased natural variability and imperfect co-locations or differences in smoothing. Model data can help to reduce these, e.g. Verhoelst et al. (2015) showed recently that MACC (IFS-MOZART) and MERRA reanalysis fields allow them to close the error budget for total ozone column validation studies. However, further work is needed in the context of vertical profile validation.

It is even more challenging to verify the GCOS requirements for stability. Figure 12 (right panels) shows that the verification of a 1 % decade^{-1} target with 95 % confidence is possible for just a few records (SAGE II, Aura MLS) and only in the middle stratosphere. In general, the analysis is not sensitive to network-averaged drifts below 2–3 % decade^{-1} in the middle stratosphere. In the upper and lower stratosphere, focus regions for current trend studies, the 2σ uncertainty on the drift is 3–4 % decade^{-1} or worse. In addition, a ground-based assessment of the meridional structure of satellite drift is currently infeasible. This is due to a lack of stations (with a long data record) in certain latitude bands and the considerable observed scatter in the single-station drift estimates. However, there is some room for improvement. The best sampled comparison time series yield 1σ drift uncertainties as low as 0.7 % decade^{-1} at individual sites. But the dominant contribution to the network-averaged drift uncertainty of some recent satellite records comes from the scatter in the drift estimates across individual sites (Fig. 3). More homogeneity across the network will surely be beneficial, and this is one of the aims of the Ozonesonde Data Quality Assessment initiative (O3S-DQA). New correction schemes are being developed for the few percent biases introduced by (station- and time-dependent) changes in instrumental and post-processing set-ups, which may, ultimately, lead to more homogeneous sonde time series in time and space (Smit et al., 2012; Tarasick et al., 2016; Van Malderen et al., 2016). When successful, this may perhaps also allow an exploration of meridional drift structure. Longer time series will also help, but not to the full extent of what is actually desired. And finally, with the help of current models part of the comparison spread could be removed statistically, which should, at least in the UTLS, lead to reduced drift uncertainties. Nevertheless, we consider it improbable that in the next few years sufficient progress can be made to demonstrate that single satellite records are stable within 1 % decade^{-1} relative to ground-based network observations. At the moment, 2–3 % decade^{-1} seems a more realistic target.

8.2 Implications for merging schemes

Space-based instruments are rarely operational for much more than a decade. Various groups have therefore produced multi-decade data sets from a series of individual records. The longest record, spanning 42 years, is based on measurements by nine SBUV nadir-viewing instruments (Bhartia et al., 2013) and was validated by Kramarova et al. (2013). Merged records based on limb/occultation instruments include SAGE-GOMOS (Kyrölä et al., 2013), SAGE-OSIRIS (Bourassa et al., 2014; Sioris et al., 2014), GOZCARDS (Froidevaux et al., 2015), SWOOSH (**?**) and Ozone_cci (Sofieva et al., 2013), all listed in Table 6. These Level-3 data are typically reported as monthly averaged ozone over 5–10° latitude bins. A recent intercomparison by Tummon et al. (2015) showed that the differences between the merged limb/occultation data sets are dominated by the differences between the underlying data sets and to a lesser extent by differences between the merging algorithms. This shows the importance of a detailed understanding of the consistency between the Level-2 records in order to understand the merged product. In addition, comprehensive intercomparison studies (such as Jones et al. (2009); Dupuy et al. (2009); Nair et al. (2012); Tegtmeier et al. (2013); Adams et al. (2014); Laeng et al. (2014) and this work), can guide the design of the merging algorithms so as to reduce the impact of unfavorable Level-2 characteristics.

Although it is well known that the bias correction scheme should be altitude-latitude dependent, further improvements could be made. The inclusion of a diurnal and seasonal component may be pertinent, as we found sunrise-sunset bias differences for a few solar occultation instruments and a pronounced seasonal dependence of the bias and short-term variability of e.g. Arctic SCIAMACHY data. We also reported that the single Level-2 profile noise of SMR and SCIAMACHY is considerably higher than that of other records. Averaged profiles will be sufficiently precise over large bins (monthly, 5° latitude) since both instruments are dense samplers, but this may not be the case at finer spatio-temporal resolutions. Our assessment of stability furthermore demonstrates the potential of drift correction schemes, especially when HALOE, OSIRIS or SCIAMACHY data are involved (and likely GOMOS and SMR as well). Eckert et al. (2014) have recently explored this approach, by correcting MIPAS trends for a drift relative to Aura MLS. In practice, however, the drift estimate between two satellite records is not sufficiently well constrained, especially for a short overlap period, which makes it very challenging to obtain robust corrections. Finally, the impact of the auxiliary data should not be forgotten, since profile representation conversions are typically required. We observed considerable changes in bias (MIPAS, SCIAMACHY and, to a lesser ex-

tent, UARS/Aura MLS) and stability (UARS/Aura MLS) due to the auxiliary data provided along with the ozone data sets. The use of a common source of stable auxiliary profiles eliminates additional discrepancies between the contributing records. Our results suggest that ECMWF (operational and ERA-Interim) and MERRA fields impact ozone trends in a consistent way over the entire stratosphere.

8.3 Are observed trend differences due to drift?

Recently, a number of regression analyses were carried out on gridded ozone profile data from a variety of limb and occultation instruments. A few studies considered single records (Eckert et al., 2014; Gebhardt et al., 2014), others a combination of two (Kyrölä et al., 2013; Laine et al., 2014; Bourassa et al., 2014; Sioris et al., 2014) or more data sets (Tummon et al., 2015; WMO, 2014; Harris et al., 2015). The resulting profile trends are generally in reasonable agreement, but notable differences are observed in some parts of the stratosphere. The SCIAMACHY data set retrieved by the IUP Bremen Level-2 processor, for instance, suggests a 2004–2012 trend in the Tropics around 35 km that is 4–6 % decade^{-1} more negative than OSIRIS and Aura MLS data (Gebhardt et al., 2014). A combined SAGE-OSIRIS record, on the other hand, produces more positive post-1998 trends in the uppermost stratosphere, by 3–4 % decade^{-1} at mid northern latitudes (Bourassa et al., 2014; Tummon et al., 2015; Harris et al., 2015). Two records that combine SAGE and GOMOS data lead to considerably more negative trends than other data sets in the lower stratosphere (Tummon et al., 2015; Harris et al., 2015).

Many of the ozone trend differences cannot be explained by statistical uncertainty. Our ground-based assessment of decadal stability suggests that these may be interpreted, at least for the better part, in terms of instrumental drift. Indeed, we noted a $+8$ % decade^{-1} drift above 40 km for OSIRIS and a -9 % decade^{-1} drift for SCIAMACHY[4] around 35 km. Additionally, we found indications of a -5 % decade^{-1} drift of GOMOS below 20 km. These quite successful interpretations of some recent ozone trend differences builds additional confidence in our single-instrument drift estimates, which could therefore be employed as 1σ systematic uncertainty for long-term trend results for the corresponding records.

No studies have been performed so far of the decadal stability of the merged data sets. Yet, there is also a clear need for realistic drift estimates for such data sets (Harris et al., 2015). We therefore make a first attempt to provide these for the merged records used by the recent WMO and SI2N assessments (WMO, 2014; Harris et al., 2015). Table 6 presents drift estimates for three stratospheric layers and for two time periods typically differentiated in trend analyses. Before 1997 all merged records rely on SAGE II observations, which are stable to within 1–1.5 % decade^{-1} depending on altitude. Since GOZCARDS and SWOOSH include HALOE data, the drift is possibly somewhat larger in the middle stratosphere. Producing post-1998 estimates is a more intricate problem, due to the increasing number of contributing instruments, and due to the fact that none of these cover the entire period. We are inclined to a conservative approach, giving figures that should be considered upper limits to the actual drift. The SAGE-GOMOS record will be impacted by negative GOMOS drifts in the lower stratosphere. The SAGE-OSIRIS trends should be considered more uncertain in the upper stratosphere due to drifting OSIRIS data. Records that use Aura MLS as backbone (GOZCARDS, SWOOSH) should not be more unstable than about 2 % decade^{-1} in the stratosphere. A merged Ozone_cci data set is likely also prone to larger uncertainty in the upper stratosphere (drifting OSIRIS, SCIAMACHY, SMR) and to some extent in the lower stratosphere as well (GOMOS). We stress that a more rigorous assessment is needed, since the estimates in Table 6 may well overestimate the actual drift. This work is currently on-going, following an approach similar to that by e.g. Mears et al. (2011) and Frith et al. (2014).

[4]Our analysis is based on the operational SCIAMACHY SGP v5.02 data set. The v2.9 data record by the IUP Bremen Level-2 processor drifts by -5 % decade^{-1} between 30–35 km (Lambert et al., 2014)

9 Conclusions

Ground-based network observations by ozonesonde and stratospheric lidar instruments allowed us to assess the quality of fourteen records of the vertical distribution of ozone, collected by limb and occultation instruments over the past three decades. We considered three aspects of satellite data quality: the stability at decadal time scale (or drift), the overall bias, and the short-term variability. Further investigation of the vertical, meridional and seasonal structure of these parameters, together with their dependence on auxiliary data, revealed common and distinguishing features between satellite instruments. Such a comprehensive analysis serves two main objectives. First, to verify whether the spatio-temporal patterns of atmospheric ozone are correctly reproduced by the individual instruments at different scales. Second, to assess the consistency between satellite records, which is vital for their synergistic exploitation, a topic that has received increased interest in recent years.

We start our concluding remarks by distilling the general tendencies, saving some prominent exceptions for the following paragraph. Typically, we found a satellite bias better than ± 5 % between 20–40 km ($\sim$ 2–50 hPa), increasing slowly towards the stratopause (± 10 %) and quite rapidly towards the tropopause (± 15 % and more). A similar vertical dependence was observed for the comparison spread. It generally ranges from 5 to 12 % between 20–40 km and increases towards the stratopause (15–20 %) and tropopause (40 % and more). The precision of the records is actually

better than suggested by the observed spread in the comparisons, since the latter also includes the precision of the ground-based record and, especially in the UTLS, the random uncertainties due to differences in co-location and horizontal smoothing. Nevertheless, the altitude at which the quality of UTLS observations starts to degrade rapidly is clearly not only determined by the tropopause. It also depends on the measurement technique and instrument (e.g. UTLS observations of UV-visible star occultations being less sensitive than those of infrared emissions at the limb). There were furthermore no evident signs of seasonal patterns, except for the Arctic SCIAMACHY data which exhibit a 10 % increase in bias and spread in boreal winter and a 10 % decrease in bias and spread in boreal summer. We found no significant drifts at decadal time scales, most records are stable within about ± 5 % decade^{-1} in the middle and upper stratosphere, and, for some records, even within ± 3 % decade^{-1} (SAGE II, Aura MLS and MIPAS). However, the drift uncertainty should not be neglected, as our analysis is typically not sensitive (at 2σ) to drifts smaller than 2–3 % decade^{-1} in the middle stratosphere and 3–4 % decade^{-1} at lower and higher altitudes. The pressure and/or temperature data that accompany the satellite ozone data sets are generally well suited for the conversion between ozone units or vertical coordinates. Bias, spread and drift in non-native ozone profile representations differ, respectively, not more than about 1 %, 1 % and 1 % decade^{-1}, and somewhat more in the uppermost stratosphere.

There are of course exceptions to these general observations. We noted more pronounced biases (~ 10 %) over much of the stratosphere for POAM II and SCIAMACHY, the latter also exhibits a clear hemispheric asymmetry. Two records show markedly poorer single-profile precision: SMR (entire atmosphere) and SCIAMACHY (upper stratosphere and Arctic). And three records drift significantly: HALOE in the middle stratosphere (-5 % decade^{-1}), and, in the upper stratosphere, SCIAMACHY (-9 % decade^{-1}) and OSIRIS ($+8$ % decade^{-1}). There are also indications of a -5 % decade^{-1} or more drift in the lower stratosphere for GOMOS, and in the upper stratosphere for SMR. Further confirmation is needed however for the latter two data sets. In the meantime, we advise caution when using GOMOS and SMR measurements at these altitudes. Finally, we observed for a few records a considerable impact of the accompanying auxiliary data (e.g. GPH retrievals) on ozone quality in non-native profile representations. The ozone bias changes by 3–5 % for MIPAS and SCIAMACHY; both MLS records (UARS and Aura) show a dependence of the drift (by 3 % decade^{-1} or more) on vertical coordinate and/or ozone unit, and perhaps of the overall bias as well. We stress that these representation-dependent quality issues are unrelated to the satellite ozone retrievals themselves, and can be avoided by using another, external source of auxiliary information for any necessary conversions.

Overall, the observing system of limb and occultation instruments produces ozone profiles that meet the ~ 10–15 % accuracy requirements by climate users, most certainly over 20–40 km, and perhaps also in the lower stratosphere. However, it remains unclear whether the current Level-2 records comply with the 1–3 % decade^{-1} target on decadal stability. The combination of different data sets has received widespread interest in recent years, but also poses several challenges. Our results show that the merging schemes should be sufficiently refined to temper additional artefacts in the Level-3 data sets. Even then, the characteristics of merged records remain mostly defined by those of their contributors (Tummon et al., 2015). Multi-instrument comparison studies are therefore crucial to establish observational evidence. Indeed, we could relate the most notable differences between recent ozone profile trend studies to instrumental drift (WMO, 2014; Harris et al., 2015). This led us to a conservative estimate of the decadal stability of several merged records, which, until more rigourous analyses are performed, provides essential information for the recent trend assessments by WMO and SI2N.

Covering most limb and occultation ozone profilers of the past three decades, the ground-based networks of sonde and lidar instruments, and all major data quality indicators, this assessment is arguably the most comprehensive ground-based analysis so far. While bias and short-term variability of satellite records are well documented in the literature, this is much less the case for their long-term stability, the impact of auxiliary data and their mutual consistency. We therefore believe that this work will contribute to an improved interpretation of observation-based studies of the long-term evolution of ozone and its link to climate change. However, our results represent a snapshot of the current versions of the data sets. In the near future, improved (and for some instruments longer) ozone profile time series will be released by the satellite teams and by the ground-based observers. Their efforts may lead to more stable records, which, in turn, would increase the sensitivity to even smaller drifts. In addition, the inclusion of microwave radiometer measurements and model data should help to evaluate the stability in the mesosphere and improve current estimates in the UTLS, especially in the Tropics.

Acknowledgements. Part of this work was funded by ESA projects Multi-TASTE and VALID, by the EU under FP6 project GEOmon (FP6-2005-Global-4-036677), and by ESA's CCI Ozone project. D. Hubert, A. Keppens and T. Verhoelst acknowledge national funding from the Belgian Science Policy Office (BELSPO) and ProDEx projects SECPEA and A3C. K. Stebel acknowledges funding from the ESA/ProDEx projects Ex Val (C90190, CN1-4, 2005-2011). J. A. E. van Gijsel and D. Swart acknowledge support from the Dutch Ministry of Infrastructure and Environment. Work performed at the Jet Propulsion Laboratory was done under contract with the National Aeronautics and Space Administration. We are also grateful to C. De Clercq, D. Pieroux and S. Vandenbussche for their valu-

able input. The ozonesonde and lidar data used in this publication were obtained as part of WMO's Global Atmosphere Watch (GAW) and two of its main contributors, namely, the Network for the Detection of Atmospheric Composition Change (NDACC) and Southern Hemisphere ADditional OZonesondes programme (SHADOZ). The authors acknowledge the meticulous and sustained work of the PIs and staff at ozonesonde and lidar stations to acquire and maintain long-term ozone data records of high quality. The data records are publicly available via the NDACC Data Host Facility (http://www.ndacc.org), the SHADOZ archive (http://croc.gsfc.nasa.gov/shadoz) and the World Ozone and Ultraviolet Data Centre (http://www.woudc.org). NDACC and SHADOZ are supported by meteorological services and space agencies from many countries, with archives funded by NASA and NOAA. We acknowledge the work by the PIs of the ozonesonde (F. Posny) and stratospheric ozone lidar (T. Portafaix) observations at Réunion Island. The authors also thank the satellite science and processing teams and the contributing space agencies. Measurements from the SAGE and HALOE missions are provided and maintained through support from NASA's Earth Science Division. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency and the Natural Sciences and Engineering Research Council of Canada. SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) is a joint contribution of Germany, The Netherlands and Belgium to ESA's environmental satellite Envisat and is funded by the German (DLR) and Dutch (NIVR) space agencies with Belgian contribution via BIRA-IASB. Sweden's Odin satellite carries the atmospheric and astronomical missions OSIRIS and SMR, developed and funded jointly by the space agencies of Sweden, Canada, Finland and France. This work is dedicated to our much appreciated colleague J. Urban, who regrettably passed away.

References

Adams, C., Bourassa, A. E., Bathgate, A. F., McLinden, C. A., Lloyd, N. D., Roth, C. Z., Llewellyn, E. J., Zawodny, J. M., Flittner, D. E., Manney, G. L., Daffer, W. H., and Degenstein, D. A.: Characterization of Odin-OSIRIS ozone profiles with the SAGE II dataset, Atmos. Meas. Tech., 6, 1447–1459, doi:10.5194/amt-6-1447-2013, 2013.

Adams, C., Bourassa, A. E., Sofieva, V., Froidevaux, L., McLinden, C. A., Hubert, D., Lambert, J.-C., Sioris, C. E., and Degenstein, D. A.: Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozonesondes, Atmos. Meas. Tech., 7, 49–64, doi:10.5194/amt-7-49-2014, 2014.

Bhartia, P. K., McPeters, R. D., Flynn, L. E., Taylor, S., Kramarova, N. A., Frith, S., Fisher, B., and DeLand, M.: Solar Backscatter UV (SBUV) total ozone and profile algorithm, Atmos. Meas. Tech., 6, 2533–2548, doi:10.5194/amt-6-2533-2013, 2013.

Boone, C. D., Walker, K. A., and Bernath, P. F.: The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, chap. Version 3 Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), 103–127, A. Deepak Publishing, Hampton, Virginia, USA, http://acebox2.uwaterloo.ca/publications/2013/Version3.5retreivals2013.pdf (last access: 23 June 2015), 2013.

Bourassa, A. E., Degenstein, D. A., Randel, W. J., Zawodny, J. M., Kyrölä, E., McLinden, C. A., Sioris, C. E., and Roth, C. Z.: Trends in stratospheric ozone derived from merged SAGE II and Odin-OSIRIS satellite observations, Atmos. Chem. Phys., 14, 6983–6994, doi:10.5194/acp-14-6983-2014, 2014.

Brewer, A. W. and Milford, J. R.: The Oxford-Kew Ozone Sonde, P. R. Soc. London, 256, 470–495, doi:10.1098/rspa.1960.0120, 1960.

Calisesi, Y., Soebijanta, V. T., and van Oss, R.: Regridding of remote soundings: Formulation and application to ozone profile comparison, J. Geophys. Res., 110, 1–8, doi:10.1029/2005JD006122, 2005.

Ceccherini, S., Carli, B., and Raspollini, P.: Quality of MIPAS operational products, J. Quant. Spectrosc. Ra., 121, 45–55, doi:10.1016/j.jqsrt.2013.01.021, 2013.

Cortesi, U., Lambert, J. C., De Clercq, C., Bianchini, G., Blumenstock, T., Bracher, A., Castelli, E., Catoire, V., Chance, K. V., De Mazière, M., Demoulin, P., Godin-Beekmann, S., Jones, N., Jucks, K., Keim, C., Kerzenmacher, T., Kuellmann, H., Kuttippurath, J., Iarlori, M., Liu, G. Y., Liu, Y., McDermid, I. S., Meijer, Y. J., Mencaraglia, F., Mikuteit, S., Oelhaf, H., Piccolo, C., Pirre, M., Raspollini, P., Ravegnani, F., Reburn, W. J., Redaelli, G., Remedios, J. J., Sembhi, H., Smale, D., Steck, T., Taddei, A., Varotsos, C., Vigouroux, C., Waterfall, A., Wetzel, G., and Wood, S.: Geophysical validation of MIPAS-ENVISAT operational ozone data, Atmos. Chem. Phys., 7, 4807–4867, doi:10.5194/acp-7-4807-2007, 2007.

Croux, C., Dhaene, G., and Hoorelbeke, D.: Robust standard errors for robust estimators, http://www.econ.kuleuven.ac.be/eng/ew/discussionpapers/Dps03/Dps0316.pdf (last access: 23 June 2015), Discussion Paper Series (DPS) 03.16, 2004.

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmos. Meas. Tech., 6, 3539–3561, doi:10.5194/amt-6-3539-2013, 2013.

Damadeo, R. P., Zawodny, J. M., and Thomason, L. W.: Reevaluation of stratospheric ozone trends from SAGE II data using a simultaneous temporal and spatial analysis, Atmos. Chem. Phys., 14, 13455–13470, doi:10.5194/acp-14-13455-2014, 2014.

Danilin, M. Y., Ko, M. K. W., Froidevaux, L., Santee, M. L., Lyjak, L. V., Bevilacqua, R. M., Zawodny, J. M., Sasano, Y., Irie, H., Kondo, Y., Russell III, J. M., Scott, C. J., and Read, W. G.: Trajectory hunting as an effective technique to validate multiplatform measurements: Analysis of the MLS, HALOE, SAGE-II, ILAS, and POAM-II data in October-November 1996, J. Geophys. Res., 107, 4420, doi:10.1029/2001JD002012, 2002.

David, C., Haefele, A., Keckhut, P., Marchand, M., Jumelet, J., Leblanc, T., Cenac, C., Laqui, C., Porteneuve, J., Haeffelin, M., Courcoux, Y., Snels, M., Viterbini, M., and Quatrevalet, M.: Evaluation of stratospheric ozone, temperature, and aerosol profiles from the LOANA lidar in Antarctica, Polar Science, 6, 209–225, doi:j.polar.2012.07.001, 2012. Check ref.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D., Read, W. G., Voemel, H., Selkirk, H., Fujiwara, M., and Damadeo, R. P.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies, AMT/ACP/ESSD, submitted, 2016. Check ref.

De Clercq, C.: Implémentation du système global pour la mesure de la composition atmosphérique: concepts et méthodes pour une analyse intégrée des données satellitales, Ph.D. thesis, BIRA-IASB and Université Libre de Bruxelles, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C. an d Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Degenstein, D. A., Bourassa, A. E., Roth, C. Z., and Llewellyn, E. J.: Limb scatter ozone retrieval from 10 to 60 km using a multiplicative algebraic reconstruction technique, Atmos. Chem. Phys., 9, 6521–6529, doi:10.5194/acp-9-6521-2009, 2009.

Deniel, C., Dalaudier, F., Chassefiere, E., Bevilacqua, R., Shettle, E., Hoppel, K., Hornstein, J., Lumpe, J., Rusch, D., and Randall, C.: A comparative study of POAM II and electrochemical concentration cell ozonesonde measurements obtained over northern Europe, J. Geophys. Res., 102, 23629–23642, doi:10.1029/97JD01665, 1997.

Dupuy, E., Walker, K. A., Kar, J., Boone, C. D., McElroy, C. T., Bernath, P. F., Drummond, J. R., Skelton, R., McLeod, S. D., Hughes, R. C., Nowlan, C. R., Dufour, D. G., Zou, J., Nichitiu, F., Strong, K., Baron, P., Bevilacqua, R. M., Blumenstock, T., Bodeker, G. E., Borsdorff, T., Bourassa, A. E., Bovensmann, H., Boyd, I. S., Bracher, A., Brogniez, C., Burrows, J. P., Catoire, V., Ceccherini, S., Chabrillat, S., Christensen, T., Coffey, M. T., Cortesi, U., Davies, J., De Clercq, C., Degenstein, D. A., De Mazière, M., Demoulin, P., Dodion, J., Firanski, B., Fischer, H., Forbes, G., Froidevaux, L., Fussen, D., Gerard, P., Godin-Beekmann, S., Goutail, F., Granville, J., Griffith, D., Haley, C. S., Hannigan, J. W., Höpfner, M., Jin, J. J., Jones, A., Jones, N. B., Jucks, K., Kagawa, A., Kasai, Y., Kerzenmacher, T. E., Kleinböhl, A., Klekociuk, A. R., Kramer, I., Küllmann, H., Kuttippurath, J., Kyrölä, E., Lambert, J.-C., Livesey, N. J., Llewellyn, E. J., Lloyd, N. D., Mahieu, E., Manney, G. L., Marshall, B. T., McConnell, J. C., McCormick, M. P., McDermid, I. S., McHugh, M., McLinden, C. A., Mellqvist, J., Mizutani, K., Murayama, Y., Murtagh, D. P., Oelhaf, H., Parrish, A., Petelina, S. V., Piccolo, C., Pommereau, J.-P., Randall, C. E., Robert, C., Roth, C., Schneider, M., Senten, C., Steck, T., Strandberg, A., Strawbridge, K. B., Sussmann, R., Swart, D. P. J., Tarasick, D. W., Taylor, J. R., Tétard, C., Thomason, L. W., Thompson, A. M., Tully, M. B., Urban, J., Vanhellemont, F., Vigouroux, C., von Clarmann, T., von der Gathen, P., von Savigny, C., Waters, J. W., Witte, J. C., Wolff, M., and Zawodny, J. M.: Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), Atmos. Chem. Phys., 9, 287–343, doi:10.5194/acp-9-287-2009, 2009.

Eckert, E., von Clarmann, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Leblanc, T., McDermid, S., Pastel, M., Steinbrecht, W., Swart, D. P. J., Walker, K. A., and Bernath, P. F.: Drift-corrected trends and periodic variations in MIPAS-IMK/IAA ozone measurements, Atmos. Chem. Phys., 14, 2571–2589, doi:10.5194/acp-14-2571-2014, 2014.

Efron, B. and Tibshirani, R.: Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, Statist. Sci., 1, 54–75, doi:10.1214/ss/1177013815, 1986.

ESA: GOMOS Level 2 processor version GOMOS/6.01 Readme, Tech. rep. ENVI-GSOP-EOGD-QD-12-0117, https://earth.esa.int/documents/700255/708000/RMF_0117_GOM_NL__2P_Disclaimers.pdf (last access: 23 June 2015), 2012a.

ESA: MIPAS Level 2 ML2PP Version 6 Readme, Tech. rep. ENVI-GSOP-EOGD-QD-12-0116, https://earth.esa.int/documents/700255/707722/MIP_NL_2P_README_V6.0.pdf (last access: 23 June 2015), 2012b.

ESA: Readme file for SCIAMACHY Level 2 version 5.02 products, Tech. rep. ENVI-GSOP-EOGD-QD-13-0118, https://earth.esa.int/documents/700255/708683/SCIAMACHY+L2+Quality+Readme+file (last access: 23 June 2015), 2013.

Frith, S. M., Kramarova, N. A., Stolarski, R. S., McPeters, R. D., Bhartia, P. K., and Labow, G. J.: Recent changes in total column ozone based on the SBUV Version 8.6 Merged Ozone Data Set, J. Geophys. Res., 119, 9735–9751, doi:10.1002/2014JD021889, 2014.

Froidevaux, L., Anderson, J., Wang, H.-J., Fuller, R. A., Schwartz, M. J., Santee, M. L., Livesey, N. J., Pumphrey, H. C., Bernath, P. F., Russell III, J. M., and McCormick, M. P.: Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, Atmos. Chem. Phys., 15, 10471–10507, doi:10.5194/acp-15-10471-2015, 2015.

GCOS: Systematic Observation Requirements for Satellite-based Products for Climate, 2011 update, p. 127, http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf (last access: 23 June 2015), 2011.

Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., Bovensmann, H., Burrows, J. P., Degenstein, D., Froidevaux, L., and Thompson, A. M.: Stratospheric ozone trends and variability as seen by SCIAMACHY from 2002 to 2012, Atmos. Chem. Phys., 14, 831–846, doi:10.5194/acp-14-831-2014, 2014.

Godin, S., Carswell, A. I., Donovan, D. P., Claude, H., Steinbrecht, W., McDermid, I. S., McGee, T. J., Gross, M. R., Nakane, H., Swart, D. P. J., Bergwerff, H. B., Uchino, O., von der Gathen, P., and Neuber, R.: Ozone differential absorption lidar algorithm intercomparison, Appl. Opt., 38, 6225–6236, doi:10.1364/AO.38.006225, 1999.

Godin, S., Bergeret, V., Bekki, S., David, C., and Mégie, G.: Study of the interannual ozone loss and the permeability of the Antarctic polar vortex from aerosol and ozone lidar measurements in Dumont d'Urville (66.4S, 140E), J. Geophys. Res. Atmos., 106, 1311–1330, doi:10.1029/2000JD900459, 2001.

Harris, N. R. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C. D., Bourassa, A., Davis, S. M., Degenstein, D., Delcloo, A., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Jones, N., Kurylo, M. J., Kyrölä, E., Laine, M., Leblanc, S. T., Lambert, J.-C., Liley, B., Mahieu, E., Maycock, A., de Mazière, M., Parrish, A., Querel, R., Rosenlof, K. H., Roth, C., Sioris, C., Staehelin, J., Stolarski, R. S., Stübi, R., Tamminen, J., Vigouroux, C., Walker, K., Wang, H. J., Wild, J., and Zawodny, J. M.: Past changes in the vertical distribution of ozone, Part 3:

Analysis and interpretation of trends, Atmos. Chem. Phys., 15, 9965–9982, doi:10.5194/acp-15-9965-2015, 2015.

Hassler, B., Petropavlovskikh, I., Staehelin, J., August, T., Bhartia, P. K., Clerbaux, C., Degenstein, D., Mazière, M. D., Dinelli, B. M., Dudhia, A., Dufour, G., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Granville, J., Harris, N. R. P., Hoppel, K., Hubert, D., Kasai, Y., Kurylo, M. J., Kyrölä, E., Lambert, J.-C., Levelt, P. F., McElroy, C. T., McPeters, R. D., Munro, R., Nakajima, H., Parrish, A., Raspollini, P., Remsberg, E. E., Rosenlof, K. H., Rozanov, A., Sano, T., Sasano, Y., Shiotani, M., Smit, H. G. J., Stiller, G., Tamminen, J., Tarasick, D. W., Urban, J., van der A, R. J., Veefkind, J. P., Vigouroux, C., von Clarmann, T., von Savigny, C., Walker, K. A., Weber, M., Wild, J., and Zawodny, J. M.: Past changes in the vertical distribution of ozone – Part 1: Measurement techniques, uncertainties and availability, Atmos. Meas. Tech., 7, 1395–1427, doi:10.5194/amt-7-1395-2014, 2014.

Hervig, M. and McHugh, M.: Cirrus detection using HALOE measurements, Geophys. Res. Lett., 26, 719–722, doi:10.1029/1999GL900069, 1999.

Inai, Y., Shiotani, M., Fujiwara, M., Hasebe, F., and Vömel, H.: Altitude misestimation caused by the Vaisala RS80 pressure bias and its impact on meteorological profiles, Atmos. Meas. Tech., 8, 4043–4054, doi:10.5194/amt-8-4043-2015, 2015.

Jégou, F., Urban, J., de La Noë, J., Ricaud, P., Le Flochmoën, E., Murtagh, D. P., Eriksson, P., Jones, A., Petelina, S., Llewellyn, E. J., Lloyd, N. D., Haley, C., Lumpe, J., Randall, C., Bevilacqua, R. M., Catoire, V., Huret, N., Berthet, G., Renard, J. B., Strong, K., Davies, J., Mc Elroy, C. T., Goutail, F., and Pommereau, J. P.: Technical Note: Validation of Odin/SMR limb observations of ozone, comparisons with OSIRIS, POAM III, ground-based and balloon-borne instruments, Atmos. Chem. Phys., 8, 3385–3409, doi:10.5194/acp-8-3385-2008, 2008.

Joint Committee for Guides in Metrology: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, Tech. rep., http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf (last access: 23 June 2015), 2012.

Jones, A., Murtagh, D., Urban, J., Eriksson, P., and Rösevall, J.: Intercomparison of Odin/SMR ozone measurements with MIPAS and balloon sonde data, Can. J. Phys., 85, 1111–1123, doi:10.1139/p07-118, 2007.

Jones, A., Urban, J., Murtagh, D. P., Eriksson, P., Brohede, S., Haley, C., Degenstein, D., Bourassa, A., von Savigny, C., Sonkaew, T., Rozanov, A., Bovensmann, H., and Burrows, J.: Evolution of stratospheric ozone and water vapour time series studied with satellite measurements, Atmos. Chem. Phys., 9, 6055–6075, doi:10.5194/acp-9-6055-2009, 2009.

Kar, J., McElroy, C., Drummond, J., Zou, J., Nichitiu, F., Walker, K., Randall, C., Nowlan, C., Dufour, D., Boone, C., Bernath, P., Trepte, C., Thomason, L., and McLinden, C.: Initial comparison of ozone and NO_2 profiles from ACE-MAESTRO with balloon and satellite data, J. Geophys. Res., 112, 1–14, doi:10.1029/2006JD008242, 2007.

Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., Barnes, J., Baray, J.-L., Bencherif, H., Claude, H., di Sarra, A., Fiocco, G., Hansen, G., Hauchecorne, A., Leblanc, T., Lee, C., Pal, S., Megie, G., Nakane, H., Neuber, R., Steinbrecht, W., and Thayer, J.: Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change, J. Environ. Monit., 6, 721–733, doi:10.1039/b404256e, 2004.

Kobayashi, J. and Toyama, Y.: On various methods of measuring the vertical distribution of atmospheric ozone (III) – Carbon iodine type chemical ozonesonde, Pap. Met. Geophys., 17, 113–126, https://www.jstage.jst.go.jp/article/mripapers1950/17/2/17_97/_article/references, 1966.

Komhyr, W. D.: Electrochemical concentration cells for gas analysis, Ann. Geophys., 25, 203–210, 1969.

Kramarova, N. A., Frith, S. M., Bhartia, P. K., McPeters, R. D., Taylor, S. L., Fisher, B. L., Labow, G. J., and DeLand, M. T.: Validation of ozone monthly zonal mean profiles obtained from the version 8.6 Solar Backscatter Ultraviolet algorithm, Atmos. Chem. Phys., 13, 6887–6905, doi:10.5194/acp-13-6887-2013, 2013.

Krijger, J. M., Snel, R., van Harten, G., Rietjens, J. H. H., and Aben, I.: Mirror contamination in space I: mirror modelling, Atmos. Meas. Tech., 7, 3387–3398, doi:10.5194/amt-7-3387-2014, 2014.

Kyrölä, E., Tamminen, J., Sofieva, V., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton d'Andon, O., Barrot, G., Guirlet, M., Mangin, A., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: Retrieval of atmospheric parameters from GOMOS data, Atmos. Chem. Phys., 10, 11881–11903, doi:10.5194/acp-10-11881-2010, 2010.

Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, S.-M., Tukiainen, S., Zawodny, J., and Thomason, L.: Combined SAGE II-GOMOS ozone profile data set for 1984-2011 and trend analysis of the vertical distribution of ozone, Atmos. Chem. Phys., 13, 10645–10658, doi:10.5194/acp-13-10645-2013, 2013.

Laeng, A., Grabowski, U., von Clarmann, T., Stiller, G., Glatthor, N., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., Sofieva, V., Petropavlovskikh, I., Hubert, D., Bathgate, T., Bernath, P., Boone, C. D., Clerbaux, C., Coheur, P., Damadeo, R., Degenstein, D., Frith, S., Froidevaux, L., Gille, J., Hoppel, K., McHugh, M., Kasai, Y., Lumpe, J., Rahpoe, N., Toon, G., Sano, T., Suzuki, M., Tamminen, J., Urban, J., Walker, K., Weber, M., and Zawodny, J.: Validation of MIPAS IMK/IAA V5R_O3_224 ozone profiles, Atmos. Meas. Tech., 7, 3971–3987, doi:10.5194/amt-7-3971-2014, 2014.

Laeng, A., Hubert, D., Verhoelst, T., von Clarmann, T., Dinelli, B., Dudhia, A., Raspollini, P., Stiller, G., Grabowski, U., Keppens, A., Kiefer, M., Sofieva, V., Froidevaux, L., Walker, K., Lambert, J.-C., and Zehner, C.: The ozone climate change initiative: Comparison of four Level-2 processors for the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), Remote Sens. Environ., 162, 316–343, doi:10.1016/j.rse.2014.12.013, 2015.

Laine, M., Latva-Pukkila, N., and Kyrölä, E.: Analysing time-varying trends in stratospheric ozone time series using the state space approach, Atmos. Chem. Phys., 14, 9707–9725, doi:10.5194/acp-14-9707-2014, 2014.

Lambert, J.-C., Balis, D., Delcloo, A., Goutail, F., Granville, J., Hubert, D., Keppens, A., Kivi, R., Koukouli, ML., Pommereau, J.-P., Stübi, R., Verhoelst, T., Loyola, D., Siddans, R., van der A, R. J., Laeng, A., Sofieva, V., Weber, W., Van Roozendael, M., and Zehner, C.: Ozone_cci – Product Validation and Intercomparison Report (PVIR), Tech. rep., BIRA-IASB, http:

//www.esa-ozone-cci.org/?q=webfm_send/148 (last access: 23 June 2015), 2014.

Lambert, J.-C. et al.: Past changes in the Vertical Distribution of Ozone, Part II: Measurement Intercomparisons, Atmos. Meas. Tech. Disc., in preparation, 2016.

Lichtenberg, G.: SCIAMACHY Offline Processor Level1b-2 ATBD – Algorithm Theoretical Baseline Document (SGP OL Version 5), Tech. rep. ENV-ATB-QWG-SCIA-0085, DLR-IMF, 2011.

Livesey, N. J., Read, W. G., Froidevaux, L., Waters, J. W., Santee, M. L., Pumphrey, H. C., Wu, D. L., Shippony, Z., and Jarnot, R. F.: The UARS Microwave Limb Sounder version 5 data set: Theory, characterization, and validation, J. Geophys. Res., 108, 4378, doi:10.1029/2002JD002273, 2003.

Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu., D. L.: Aura Microwave Limb Sounder (MLS). Version 3.3 Level 2 data quality and description document, Tech. rep., Jet Propulsion Laboratory, 2011.

Livesey, N. J., Logan, J. A., Santee, M. L., Waters, J. W., Doherty, R. M., Read, W. G., Froidevaux, L., and Jiang, J. H.: Interrelated variations of O_3, CO and deep convection in the tropical/subtropical upper troposphere observed by the Aura Microwave Limb Sounder (MLS) during 2004-2011, Atmos. Chem. Phys., 13, 579–598, doi:10.5194/acp-13-579-2013, 2013a.

Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu., D. L.: Aura Microwave Limb Sounder (MLS). Version 3.3 and 3.4 Level 2 data quality and description document, Tech. rep., Jet Propulsion Laboratory, http://mls.jpl.nasa.gov/data/v3_data_quality_document.pdf (last access: 23 June 2015), 2013b.

Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L., Mill'an Valle, L. F., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., and Martin, E.: Aura Microwave Limb Sounder (MLS). Version 4.2x Level 2 data quality and description document, Tech. rep., Jet Propulsion Laboratory, http://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf (last access: 23 June 2015), 2015.

Lumpe, J. D., Bevilacqua, R. M., Hoppel, K. W., Krigman, S. S., Kriebel, D. L., Debrestian, D. J., Randall, C. E., Rusch, D. W., Brogniez, C., Ramananahérisoa, R., Shettle, E. P., Olivero, J. J., Lenoble, J., and Pruvost, P.: POAM II retrieval algorithm and error analysis, J. Geophys. Res., 102, 23593–23614, doi:10.1029/97JD00906, 1997.

Lumpe, J. D., Bevilacqua, R., Hoppel, K., and Randall, C.: POAM III retrieval algorithm and error analysis, J. Geophys. Res., 107, 4575, doi:10.1029/2002JD002137, 2002.

McDermid, I. S., Godin, S. M., and Lindqvist, L. O.: Ground-based laser DIAL system for long-term measurements of stratospheric ozone, Appl. Opt., 29, 3603–3612, doi:10.1364/AO.29.003603, 1990.

McElroy, C., Nowlan, C., Drummond, J., Bernath, P., Barton, D., Dufour, D., Midwinter, C., Hall, R., Ogyu, A., Ullberg, A., Wardle, D., Kar, J., Zou, J., Nichitiu, F., Boone, C., Walker, K., and Rowlands, N.: The ACE-MAESTRO instrument on SCISAT: description, performance, and preliminary results, Appl. Opt., 46, 4341–4356, doi:10.1364/AO.46.004341, 2007.

McGee, T. J., Whiteman, D. N., Ferrare, R. A., Butler, J. J., and Burris, J. F.: STROZ LITE: Stratospheric Ozone Lidar Trailer Experiment, Opt. Eng., 30, 31–39, doi:10.1117/12.55771, 1991.

McLinden, C. A. and Fioletov, V.: Quantifying stratospheric ozone trends: Complications due to stratospheric cooling, Geophys. Res. Lett., 38, L03808, doi:10.1029/2010GL046012, 2011.

Mears, C. A., Wentz, F. J., Thorne, P., and Bernie, D.: Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique, J. Geophys. Res., 116, doi:10.1029/2010jd014954, 2011.

Mégie, G., Allain, J. Y., Chanin, M. L., and Blamont, J. E.: Vertical profile of stratospheric ozone by lidar sounding from the ground, Nature, 270, 329–331, doi:10.1038/270329a0, 1977.

Mieruch, S., Weber, M., von Savigny, C., Rozanov, A., Bovensmann, H., Burrows, J. P., Bernath, P. F., Boone, C. D., Froidevaux, L., Gordley, L. L., Mlynczak, M. G., Russell III, J. M., Thomason, L. W., Walker, K. A., and Zawodny, J. M.: Global and long-term comparison of SCIAMACHY limb ozone profiles with correlative satellite data (2002-2008), Atmos. Meas. Tech., 5, 771–788, doi:10.5194/amt-5-771-2012, 2012.

Morris, G., Gleason, J., Russell III, J., Schoeberl, M., and McCormick, M.: A comparison of HALOE V19 with SAGE II V6.00 ozone observations using trajectory mapping, J. Geophys. Res., 107, 4177, doi:10.1029/2001JD000847, 2002.

Muhlbauer, A., Spichtinger, P., and Lohmann, U.: Application and Comparison of Robust Linear Regression Methods for Trend Estimation, J. Appl. Meteorol. Clim., 48, 1961–1970, doi:10.1175/2009JAMC1851.1, 2009.

Nair, P. J., Godin-Beekmann, S., Pazmiño, A., Hauchecorne, A., Ancellet, G., Petropavlovskikh, I., Flynn, L. E., and Froidevaux, L.: Coherence of long-term stratospheric ozone vertical distribution time series used for the study of ozone recovery at a northern mid-latitude station, Atmos. Chem. Phys., 11, 4957–4975, doi:10.5194/acp-11-4957-2011, 2011.

Nair, P. J., Godin-Beekmann, S., Froidevaux, L., Flynn, L. E., Zawodny, J. M., Russell III, J. M., Pazmiño, A., Ancellet, G., Steinbrecht, W., Claude, H., Leblanc, T., McDermid, S., van Gijsel, J. A. E., Johnson, B., Thomas, A., Hubert, D., Lambert, J.-C., Nakane, H., and Swart, D. P. J.: Relative drifts and stability of satellite and ground-based stratospheric ozone profiles at NDACC lidar stations, Atmos. Meas. Tech., 5, 1301–1318, doi:10.5194/amt-5-1301-2012, 2012.

Naval Research Lab: Description of POAM III Version 4 Retrievals, https://eosweb.larc.nasa.gov/sites/default/files/project/poam3/readme/poam3_ver4_documentation.pdf (last access: 23 June 2015), 2005.

Nazaryan, H., McCormick, M. P., and Russell III, J. M.: New studies of SAGE II and HALOE ozone profile and long-term change comparisons, J. Geophys. Res., 110, 1–13, doi:10.1029/2004JD005425, 2005.

Oman, L. D., Plummer, D. A., Waugh, D. W., Austin, J., Scinocca, J. F., Douglass, A. R., Salawitch, R. J., Canty, T., Akiyoshi, H., Bekki, S., Braesicke, P., Butchart, N., Chipperfield, M. P., Cugnet, D., Dhomse, S., Eyring, V., Frith, S., Hardiman, S. C., Kinnison, D. E., Lamarque, J.-F., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Nielsen, J. E., Olivié, D., Pitari, G., Pyle, J., Rozanov, E., Shepherd, T. G., Shi-

bata, K., Stolarski, R. S., Teyssèdre, H., Tian, W., Yamashita, Y., and Ziemke, J. R.: Multimodel assessment of the factors driving stratospheric ozone evolution over the 21st century, J. Geophys. Res. Atmos., 115, n/a–n/a, doi:10.1029/2010JD014362, D24306, 2010.

Parrish, A., Boyd, I. S., Nedoluha, G. E., Bhartia, P. K., Frith, S. M., Kramarova, N. A., Connor, B. J., Bodeker, G. E., Froidevaux, L., Shiotani, M., and Sakazaki, T.: Diurnal variations of stratospheric ozone measured by ground-based microwave remote sensing at the Mauna Loa NDACC site: measurement validation and GEOSCCM model comparison, Atmos. Chem. Phys., 14, 7255–7272, doi:10.5194/acp-14-7255-2014, 2014.

Randall, C. E., Rusch, D. W., Bevilacqua, R. M., Hoppel, K. W., Lumpe, J. D., Shettle, E., Thompson, E., Deaver, L., Zawodny, J., Kyrö, E., Johnson, B., Kelder, H., Dorokhov, V. M., König-Langlo, G., and Gil, M.: Validation of POAM III ozone: Comparison with ozonesonde and satellite data, J. Geophys. Res., 108, 4367, doi:10.1029/2002JD002944, 2003.

Raspollini, P., Carli, B., Carlotti, M., Ceccherini, S., Dehn, A., Dinelli, B. M., Dudhia, A., Flaud, J.-M., López-Puertas, M., Niro, F., Remedios, J. J., Ridolfi, M., Sembhi, H., Sgheri, L., and von Clarmann, T.: Ten years of MIPAS measurements with ESA Level 2 processor V6 – Part 1: Retrieval algorithm and diagnostics of the products, Atmos. Meas. Tech., 6, 2419–2439, doi:10.5194/amt-6-2419-2013, 2013.

Rozanov, A., Eichmann, K.-U., von Savigny, C., Bovensmann, H., Burrows, J. P., von Bargen, A., Doicu, A., Hilgers, S., Godin-Beekmann, S., Leblanc, T., and McDermid, I. S.: Comparison of the inversion algorithms applied to the ozone vertical profile retrieval from SCIAMACHY limb measurements, Atmos. Chem. Phys., 7, 4763–4779, doi:10.5194/acp-7-4763-2007, 2007.

Rusch, D., Bevilacqua, R., Randall, C., Lumpe, J., Hoppel, K., Fromm, M., Debrestian, D., Olivero, J., Hornstein, J., Guo, F., and Shettle, E.: Validation of POAM II Ozone Measurements with Coincident MLS, HALOE, and SAGE II Observations, J. Geophys. Res., 102, 23615–23627, doi:10.1029/97JD00458, 1997.

Sakazaki, T., Shiotani, M., Suzuki, M., Kinnison, D., Zawodny, J. M., McHugh, M., and Walker, K. A.: Sunset-sunrise difference in solar occultation ozone measurements (SAGE II, HALOE, and ACE-FTS) and its relationship to tidal vertical winds, Atmos. Chem. Phys., 15, 829–843, doi:10.5194/acp-15-829-2015, 2015.

Schanz, A., Hocke, K., and Kämpfer, N.: Daily ozone cycle in the stratosphere: global, regional and seasonal behaviour modelled with the Whole Atmosphere Community Climate Model, Atmos. Chem. Phys., 14, 7645–7663, doi:10.5194/acp-14-7645-2014, 2014.

Seidel, D. J., Gillett, N. P., Lanzante, J. R., Shine, K. P., and Thorne, P. W.: Stratospheric temperature trends: our evolving understanding, WIREs Climate Change, 2, 592–616, doi:10.1002/wcc.125, 2011.

Simmons, A. J., Poli, P., Dee, D. P., Berrisford, P., Hersbach, H., Kobayashi, S., and Peubey, C.: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim, Q. J. Roy. Meteor. Soc., 140, 329–353, doi:10.1002/qj.2317, 2014.

Sioris, C. E., McLinden, C. A., Fioletov, V. E., Adams, C., Zawodny, J. M., Bourassa, A. E., Roth, C. Z., and Degenstein, D. A.: Trend and variability in ozone in the tropical lower stratosphere over 2.5 solar cycles observed by SAGE II and OSIRIS, Atmos. Chem. Phys., 14, 3479–3496, doi:10.5194/acp-14-3479-2014, 2014.

Smit, H., Oltmans, S., Deshler, T., Tarasick, D., Johnson, B., Schmidlin, F., Stübi, R., and Davies, J.: SI2N/O3S-DQA Activity: Guide Lines for Homogenization of Ozone Sonde Data, Tech. rep. 2.0, http://www-das.uwyo.edu/~deshler/NDACC_O3Sondes/O3s_DQA/O3S-DQA-GuidelinesHomogenization-V2-19November2012.pdf (last access: 23 June 2015), 2012.

Smit, H. G. J. and ASOPOS-panel: Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, WMO Global Atmosphere Watch report series 201, World Meteorological Organization, http://www.wmo.int/pages/prog/arep/gaw/documents/FINAL_GAW_201_Oct_2014.pdf (last access: 23 June 2015), 2014.

Sofieva, V. F., Rahpoe, N., Tamminen, J., Kyrölä, E., Kalakoski, N., Weber, M., Rozanov, A., von Savigny, C., Laeng, A., von Clarmann, T., Stiller, G., Lossow, S., Degenstein, D., Bourassa, A., Adams, C., Roth, C., Lloyd, N., Bernath, P., Hargreaves, R. J., Urban, J., Murtagh, D., Hauchecorne, A., Dalaudier, F., van Roozendael, M., Kalb, N., and Zehner, C.: Harmonized dataset of ozone profiles from satellite limb and occultation measurements, Earth System Science Data, 5, 349–363, doi:10.5194/essd-5-349-2013, 2013.

Stauffer, R. M., Morris, G. A., Thompson, A. M., Joseph, E., Coetzee, G. J. R., and Nalli, N. R.: Propagation of radiosonde pressure sensor errors to ozonesonde measurements, Atmos. Meas. Tech., 7, 65–79, doi:10.5194/amt-7-65-2014, 2014.

Street, J. O., Carroll, R. J., and Ruppert, D.: A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares, The American Statistician, 42, 152–154, doi:10.1080/00031305.1988.10475548, 1988.

Sun, B., Reale, A., Schroeder, S., Seidel, D. J., and Ballish, B.: Toward improved corrections for radiation-induced biases in radiosonde temperature observations, J. Geophys. Res., 118, 4231–4243, doi:10.1002/jgrd.50369, 2013.

Tamminen, J., Kyrölä, E., Sofieva, V. F., Laine, M., Bertaux, J.-L., Hauchecorne, A., Dalaudier, F., Fussen, D., Vanhellemont, F., Fanton-d'Andon, O., Barrot, G., Mangin, A., Guirlet, M., Blanot, L., Fehr, T., Saavedra de Miguel, L., and Fraisse, R.: GOMOS data characterisation and error estimation, Atmos. Chem. Phys., 10, 9505–9519, doi:10.5194/acp-10-9505-2010, 2010.

Tarasick, D. W., Davies, J., Smit, H. G. J. and Oltmans, S. J.: A re-evaluated Canadian ozonesonde record: measurements of the vertical distribution of ozone over Canada from 1966 to 2013, Atmos. Meas. Tech., 9, 195–214, doi:10.5194/amt-9-195-2016, 2016.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Lingenfelser, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, H. J.: The SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, J. Geophys. Res., 118, 12229–12247, doi:10.1002/2013JD019877, 2013.

Terao, Y. and Logan, J.: Consistency of time series and trends of stratospheric ozone as seen by ozonesonde, SAGE

II, HALOE, and SBUV(/2), J. Geophys. Res., 112, 1–23, doi:10.1029/2006JD007667, 2007.

Thomason, L. W., Moore, J. R., Pitts, M. C., Zawodny, J. M., and Chiou, E. W.: An evaluation of the SAGE III version 4 aerosol extinction coefficient and water vapor data products, Atmos. Chem. Phys., 10, 2159–2173, doi:10.5194/acp-10-2159-2010, 2010.

Thompson, A. M. et al.: Southern Hemisphere Additional Ozonesondes (SHADOZ) ozone climatology (2005-2009): Tropospheric and tropical tropopause layer (TTL) profiles with comparisons to OMI-based ozone products, J. Geophys. Res., 117, D23 301, doi:10.1029/2011JD016911, 2012.

Thorne, P. W., Lanzante, J. R., Peterson, T. C., Seidel, D. J., and Shine, K. P.: Tropospheric temperature trends: history of an ongoing controversy, WIREs Climate Change, 2, 66–88, doi:10.1002/wcc.80, 2011.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res., 118, 11,847–11,862, doi:10.1002/jgrd.50874, 2013.

Tummon, F., Hassler, B., Harris, N. R. P., Staehelin, J., Steinbrecht, W., Anderson, J., Bodeker, G. E., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S. M., Froidevaux, L., Kyrölä, E., Laine, M., Long, C., Penckwitt, A. A., Sioris, C. E., Rosenlof, K. H., Roth, C., Wang, H.-J., and Wild, J.: Intercomparison of vertically resolved merged satellite ozone data sets: interannual variability and long-term trends, Atmos. Chem. Phys., 15, 3021–3043, doi:10.5194/acp-15-3021-2015, 2015.

Urban, J., Lautié, N., Le Flochmoën, E., Jiménez, C., Eriksson, P., de La Noë, J., Dupuy, E., Ekström, M., El Amraoui, L., Frisk, U., Murtagh, D., Olberg, M., and Ricaud, P.: Odin/SMR limb observations of stratospheric trace gases: Level 2 processing of ClO, N_2O, HNO_3, and O_3, J. Geophys. Res., 110, doi:10.1029/2004JD005741, 2005.

van der A, R. J. et al.: Ozone_cci – User Requirement Document (URD), Tech. rep. v2.1, KNMI, http://www.esa-ozone-cci.org/?q=webfm_send/37 (last access: 23 June 2015), 2011.

van Gijsel, J. A. E., Swart, D. P. J., Baray, J.-L., Bencherif, H., Claude, H., Fehr, T., Godin-Beekmann, S., Hansen, G. H., Keckhut, P., Leblanc, T., McDermid, I. S., Meijer, Y. J., Nakane, H., Quel, E. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Tatarov, B. I., and Wolfram, E. A.: GOMOS ozone profile validation using ground-based and balloon sonde measurements, Atmos. Chem. Phys., 10, 10473–10488, doi:10.5194/acp-10-10473-2010, 2010.

Van Malderen, R., Allaart, M. A. F., De Backer, H., Smit, H. G. J., and De Muer, D.: On instrumental errors and related correction strategies of ozonesondes: possible effect on calculated ozone trends for the nearby sites Uccle and De Bilt, Atmos. Meas. Tech. Disc., doi:10.5194/amt-2015-341, 2016.

Verhoelst, T., Granville, J., Hendrick, F., Köhler, U., Lerot, C., Pommereau, J.-P., Redondas, A., Van Roozendael, M., and Lambert, J.-C.: Metrology of ground-based satellite validation: Co-location mismatch and smoothing issues of total ozone comparisons, Atmos. Meas. Tech., 8, 5039–5062, doi:10.5194/amt-8-5039-2015, 2015.

von Clarmann, T.: Validation of remotely sensed profiles of atmospheric state variables: strategies and terminology, Atmos. Chem. Phys., 6, 4311–4320, doi:10.5194/acp-6-4311-2006, 2006.

Wang, H. J., Cunnold, D. M., Froidevaux, L., and Russell, J. M.: A reference model for middle atmosphere ozone in 1992-1993, J. Geophys. Res., 104, 21 629–21 643, doi:10.1029/1999JD900412, 1999.

Wang, H. J., Cunnold, D. M., Thomason, L. W., Zawodny, J. M., and Bodeker, G. E.: Assessment of SAGE version 6.1 ozone data quality, J. Geophys. Res., 107, 4691, doi:10.1029/2002JD002418, 2002.

Wang, H.-J., Cunnold, D. M., Trepte, C., Thomason, L. W., and Zawodny, J. M.: SAGE III solar ozone measurements: Initial results, Geophys. Res. Lett., 33, doi:10.1029/2005GL025099, 2006.

Waugh, D. W., Oman, L., Kawa, S. R., Stolarski, R. S., Pawson, S., Douglass, A. R., Newman, P. A., and Nielsen, J. E.: Impacts of climate change on stratospheric ozone recovery, Geophys. Res. Lett., 36, L03805, doi:10.1029/2008GL036223, 2009.

Waymark, C., Walker, K. A., Boone, C. D. and Bernath, P. F.: ACE-FTS version 3.0 data set: validation and data processing update, in: Proceedings of the ACVE workshop, Frascati, Italy, March 2013, Ann. Geophys., 56, doi:10.4401/ag-6339, 2013.

WMO: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project – Report No. 52, http://ozone.unep.org/Assessment_Panels/SAP/Scientific_Assessment_2010/00-SAP-2010-Assement-report.pdf (last access: 23 June 2015), 2011.

WMO: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project – Report No. 55, http://ozone.unep.org/Assessment_Panels/SAP/Scientific_Assessment_2014/Scientific_Assessment_Report_2014.pdf (last access: 23 June 2015), 2014.

Table 1. Overview of the 72 ozonesonde stations considered in this work, their location and the archive the data were taken from. Time range and profile statistics reflect the total, screened sample straddling the analysis period (10/1984–5/2013), not the co-located sample (which differs per satellite instrument). All listed stations were used in the analyses of bias and comparison spread, those indicated in the last column were also used for the drift analysis.

Station	Lat. (° N)	Lon. (° E)	Responsible institute	Data archive*	Analysis period		N_{profile}	Included in drift analysis
Alert	82.5	−62.5	MSC	WOUDC	12/1987	12/2011	1244	✓
Eureka	80.0	−85.9	MSC	WOUDC	11/1992	9/2011	1318	✓
Ny-Ålesund	78.9	11.9	AWI-NA	WOUDC	10/1990	5/2013	2224	✓
Thule	76.5	−68.7	DMI	NDACC	10/1991	1/2013	349	✓
Resolute	74.7	−95.0	MSC	WOUDC	10/1984	8/2011	1020	✓
Summit	72.3	−38.3	NOAA-ESRL	NDACC	2/2005	5/2013	427	✓
Scoresbysund	70.5	−21.9	DMI	NDACC	2/1989	5/2013	1169	✓
Sodankylä	67.4	26.6	FMI	NDACC	11/1991	12/2010	1085	✓
Edmonton	53.5	−114.1	MSC	WOUDC	10/1984	8/2011	1193	✓
Goose Bay	53.3	−60.4	MSC	WOUDC	10/1984	8/2011	1272	✓
Lindenberg	52.2	14.1	DWD-MOL	WOUDC	10/1984	5/2013	1660	✓
De Bilt	52.1	5.2	KNMI	NDACC	11/1992	12/2012	1061	✓
Vanscoy	52.0	−107.0	MSC	WOUDC	8/1990	9/2004	60	
Valentia	51.9	−10.2	ME	WOUDC	1/1994	12/2012	555	✓
Uccle	50.8	4.3	RMIB	WOUDC	10/1984	6/2012	3712	✓
Gimli	50.6	−97.0	MSC	WOUDC	7/1985	8/1985	10	
Bratt's Lake	50.2	−104.7	MSC	WOUDC	12/2003	9/2011	402	✓
Praha	50.0	14.4	CHMI-PR	WOUDC	1/1985	4/2013	1210	✓
Kelowna	49.9	−119.4	MSC	WOUDC	11/2003	8/2011	432	✓
Hohenpeißenberg	47.8	11.0	DWD-MOHp	WOUDC	10/1984	5/2013	3586	✓
Payerne	46.8	7.0	MCH	WOUDC	10/1984	12/2012	4052	✓
Pellston	45.6	−84.7	NOAA-ESRL	WOUDC	7/2004	8/2004	5	
Pietro Capofiume	44.6	11.6	AM-IMS	WOUDC	3/1991	12/1993	95	
Egbert	44.2	−79.8	MSC	WOUDC	12/2003	8/2011	373	✓
Yarmouth	43.9	−66.1	MSC	WOUDC	10/2003	8/2011	394	✓
Sofia	42.8	23.4	BNIHM	WOUDC	11/1984	12/1991	145	
Trinidad Head	40.8	−124.2	NOAA-ESRL	WOUDC	1/1999	8/2006	197	✓
Madrid	40.5	−3.7	AEMET	WOUDC	12/1994	5/2013	738	✓
Boulder	40.0	−105.2	NOAA-ESRL	NDACC	6/1991	5/2013	1097	✓
Beltsville	39.0	−76.5	Howard U	WOUDC	8/2006	8/2006	12	
Huntsville	34.7	−86.6	UAH	WOUDC	4/1999	12/2007	574	✓
Table Mountain	34.4	−117.7	NASA-JPL	WOUDC	2/2006	8/2006	35	✓
Isfahan	32.5	51.4	MDI	WOUDC	7/1995	4/2011	151	✓
Palestine	31.8	−95.7	MSC	WOUDC	5/1985	6/1985	26	

Table 1. Continued.

Station	Lat. (° N)	Lon. (° E)	Responsible institute	Data archive*	Analysis period		N_{profile}	Included in drift analysis
Houston	29.7	−95.4	Valparaiso U	WOUDC	7/2004	8/2006	62	✓
Santa Cruz	28.5	−16.3	AEMET	WOUDC	1/1996	5/2003	322	
Izaña	28.3	−16.5	AEMET	NDACC	1/1995	3/2012	976	✓
Taipei	25.0	121.5	CWBT	WOUDC	1/2000	8/2001	64	
Hilo	19.7	−155.1	NOAA-ESRL	NDACC	7/1991	6/2010	855	✓
Tecamec	19.3	−99.2	PennState U	WOUDC	3/2006	9/2006	34	
Barbados	13.2	−59.4	NOAA-ESRL	WOUDC	7/2006	8/2006	27	
Cotonou	6.2	2.2	CNRS	SHADOZ	1/2005	1/2007	97	
Paramaribo	5.8	−55.2	KNMI	NDACC	9/1999	5/2013	534	✓
Kaashidhoo	5.0	73.5	NOAA-ESRL	WOUDC	1/1999	3/1999	54	
San Cristóbal	−0.9	−89.6	NOAA-ESRL	SHADOZ/WOUDC	3/1998	10/2008	708	✓
Nairobi	−1.3	36.8	MCH	SHADOZ/WOUDC	12/1996	12/2012	1058	✓
Malindi	−3.0	40.2	U Rome-CRPSM	SHADOZ	3/1999	1/2006	191	✓
Brazzaville	−4.3	15.2	NASA-LaRC	WOUDC	4/1990	10/1992	80	
Natal	−5.8	−35.2	INPE	SHADOZ/WOUDC	3/1990	12/2010	650	✓
Watukosek	−7.5	112.6	Hokkaido U	SHADOZ	8/1999	12/2011	573	✓
Ascension Island	−8.0	−14.4	NASA-WFF	SHADOZ/WOUDC	7/1990	8/2010	1112	✓
Porto Nacional	−10.8	−48.4	NASA-LaRC	WOUDC	9/1992	10/1992	15	
Samoa	−14.2	−170.6	NOAA-ESRL	NDACC	8/1995	5/2013	663	✓
Cuiaba	−15.6	−56.1	INPE	WOUDC	9/1992	10/1992	22	
Papeete	−18.0	−149.0	NOAA-ESRL	SHADOZ/WOUDC	7/1995	12/1999	167	
Suva	−18.1	178.4	NOAA-ESRL	SHADOZ	2/1997	12/2011	727	✓
Etosha Pan	−19.2	15.9	NASA-LaRC	WOUDC	9/1992	10/1992	15	
Réunion Island	−20.9	55.5	U La Reunion	SHADOZ	1/1998	11/2012	810	✓
Irene	−25.9	28.2	SAWS	SHADOZ/WOUDC	7/1990	10/2007	581	✓
Easter Island	−27.2	−109.4	EIMO	WOUDC	8/1995	6/1997	71	
Broadmeadows	−37.7	144.9	ABM	WOUDC	2/1999	12/2012	623	✓
Laverton	−37.9	144.8	ABM	WOUDC	10/1984	2/1999	344	✓
Lauder	−45.0	169.7	NIWA	NDACC	8/1986	5/2013	1609	✓
Macquarie	−54.5	158.9	ABM	WOUDC	3/1994	12/2012	712	✓
Marambio	−64.2	−56.6	FMI-SMNA	WOUDC	11/1988	5/2013	891	✓
Mirny	−66.5	93.0	MGO	WOUDC	7/1989	12/1991	114	
Davis	−68.6	78.0	ABM	WOUDC	2/2003	12/2012	282	✓
Syowa	−69.0	39.6	JMA	WOUDC	12/1984	5/2013	1134	✓
Neumayer	−70.7	−8.3	AWI-NM	WOUDC	3/1992	5/2013	1540	✓
Novolasarevskaya	−70.8	11.9	MGO	WOUDC	5/1985	2/1991	374	
Mac Murdo	−77.8	166.6	U Wyoming	NDACC	8/1986	10/2010	817	✓
Amundsen-Scott	−90.0	−24.8	NOAA-ESRL	NDACC	11/1990	5/2013	1463	✓

* Sources: NDACC, http://www.ndacc.org; WOUDC, http://www.woudc.org; SHADOZ, http://croc.gsfc.nasa.gov/shadoz.

Table 2. Like Table 1, but for the 13 considered stratospheric ozone lidar stations.

Station	Lat. (° N)	Lon. (° E)	Responsible institute	Data archive	Analysis period		N_{profile}	Included in drift analysis
Eureka	80.0	−85.9	EC	NDACC	2/1993	3/2009	513	✓[a]
Ny-Ålesund	78.9	11.9	AWI	NDACC	11/1991	3/2011	791	✓[a]
Andøya	69.3	16.0	NILU	NDACC	12/1994	4/2011	594	
Hohenpeißenberg	47.8	11.0	DWD-MOHp	NDACC	9/1987	5/2013	2280	✓[b]
Observatoire de Haute-Provence	43.9	5.7	LATMOS-CNRS	NDACC	7/1985	5/2013	2776	✓
Toronto	43.8	−79.5	EC	NDACC	5/1991	12/1997	235	✓
Tsukuba	36.0	140.1	NIES	NDACC	8/1988	2/2010	592	
Table Mountain	34.4	−117.7	NASA-JPL	NDACC	1/1989	5/2013	1758	✓
Mauna Loa	19.5	−155.6	NASA-JPL	NDACC	7/1993	5/2013	2401	✓
Réunion Island	−20.9	55.5	U La Reunion, CNRS	NDACC	5/2000	12/2006	85	✓
Lauder	−45.0	169.7	RIVM, NIWA	NDACC	11/1994	6/2011	1030	✓
Rio Gallegos	−51.6	−69.3	OAPA-CEILAP	NDACC	8/2005	11/2010	140	✓
Dumont d'Urville	−66.7	140.0	LATMOS-CNRS	NDACC	4/1991	2/2013	678	

[a] All Arctic lidar data are discarded in the drift analysis of the SCIAMACHY record.
[b] Hohenpeißenberg is only included in the drift analysis of satellite instruments operational in 2007.

Table 3. Overview of satellite ozone profile data records. For more details on the instrument and the retrieval technique we refer to the review by Hassler et al. (2014). Some instrument teams recommend to discard a considerable part of their ozone record for long-term studies. The asterisk in the analysis period columns denotes whether the early or late part of the mission is cropped (see text).

Instrument	Level-2 data version	Analysis period		Satellite platform	Observation geometry	Spectral range O_3 retrieval	Description data set
SAGE II	v7.0	10/1984	8/2005	ERBS	solar occultation	VIS	Damadeo et al. (2013)
SAGE III	v4.0	3/2002	11/2005	METEOR-3M	solar occultation	UV-VIS	Thomason et al. (2010)
HALOE	v19	10/1991	11/2005	UARS	solar occultation	MIR	Nazaryan et al. (2005)
UARS MLS	v5	9/1991	*6/1997	UARS	limb emission	MW (205 GHz)	Livesey et al. (2003)
Aura MLS	v3.3	8/2004	5/2013	EOS-Aura	limb emission	MW (240 GHz)	Livesey et al. (2013b)
POAM II	v6	11/1993	11/1996	SPOT-3	solar occultation	VIS	Lumpe et al. (1997)
POAM III	v4	4/1998	12/2005	SPOT-4	solar occultation	VIS	Lumpe et al. (2002)
OSIRIS	v5.07	10/2001	5/2013	Odin	limb scattered	UV-VIS	Degenstein et al. (2009)
SMR	v2.1	6/2001	5/2013	Odin	limb emission	MW (501.8 GHz)	Urban et al. (2005)
GOMOS	IPF 6.01	7/2002	4/2012	Envisat	stellar occultation	UV-VIS	Kyrölä et al. (2010)
MIPAS	ML2PP 6.0	*1/2005	4/2012	Envisat	limb emission	MIR	Raspollini et al. (2013)
SCIAMACHY	SGP 5.02	8/2002	4/2012	Envisat	limb scattered	VIS	Lichtenberg (2011)
ACE-FTS	v3.0	2/2004	*9/2010	SCISAT	solar occultation	MIR	Boone et al. (2013)
MAESTRO	v1.2	2/2004	*9/2010	SCISAT	solar occultation	VIS	McElroy et al. (2007)

UV: ultraviolet; VIS: visible; MIR: mid-infrared; MW: microwave.

Table 3. Continued.

Instrument	Approximate observation time	Latitude range	Vertical range [km]	Vertical resol. [km]	Native profile representation	Source auxiliary data	Screening reference
SAGE II	sunrise & sunset	80° N – 80° S	CT–60	1	(z_{gm}, n)	MERRA (+GRAM-95): p, T	https://eosweb.larc.nasa.gov/project/sage2/sage2_release_v7_notes
SAGE III	sunset, sunrise	50–80° N, 30–50° S	6–85	1	(z_{gm}, n)	NCEP (+GRAM-95): p, T	Done by SAGE III team
HALOE	sunrise & sunset	80° N – 80° S	15–60	2.3	(p, VMR)	HALOE/NCEP: z_{gm}, T	Hervig and McHugh (1999)
UARS MLS	variable, day & night	34° N – 80° S, 34° S – 80° N	15–60	3–5	(p, VMR)	UARS MLS: z_{gp}, T	Livesey et al. (2003)
Aura MLS	01:30 & 13:30	82° N – 82° S	10–75	2.5–4	(p, VMR)	Aura MLS: z_{gp}, T	Livesey et al. (2011)
POAM II	sunrise, sunset	55–71° N, 63–88° S	15–50	1	(z_{gm}, n)	UKMO: p, T	–
POAM III	sunrise, sunset	55–71° N, 63–88° S	10–60	1–2	(z_{gm}, n)	UKMO: p, T	Naval Research Lab (2005)
OSIRIS	06:30 & 18:30	82° N – 82° S	CT–60	1–2	(z_{gm}, n)	ECMWF: p, T	Adams et al. (2013)
SMR	06:30 & 18:30	82° N – 82° S	18–60	3	(z_{gm}, VMR)	ECMWF: p, T	Jones et al. (2009)
GOMOS	22:00	90° N – 90° S	12–100	2–3	(z_{gm}, n)	ECMWF: p, T	ESA (2012a)
MIPAS	10:00 & 22:00	80° N – 80° S	6–68	3–4	(p, VMR)	MIPAS: z_{gm}, T	ESA (2012b)
SCIAMACHY	10:00	82° N – 80° S	15–40	3	(z_{gm}, n)	McLinden clim.: p, T	ESA (2013)
ACE-FTS	sunrise & sunset	85° N – 85° S	CT–95	3–4	(z_{gm}, VMR)	ACE-FTS/CMC: p/T	Dupuy et al. (2009)
MAESTRO	sunrise & sunset	85° N – 85° S	CT–100	1.5	(z_{gm}, VMR)	ACE-FTS: p, T	Kar et al. (2007)

CT stands for cloud top, p for pressure, T for temperature, $z_{\mathrm{g}m}$ for geometric altitude, $z_{\mathrm{g}p}$ for geopotential height, VMR for volume mixing ratio, n for number density.

Table 4. Overview of the drift of satellite ozone profile records relative to ozonesonde and lidar, in the lower, middle and upper stratosphere. For each altitude region we present the range of the network average of the drift ($\bar{\alpha}$) and its adjusted one-sigma uncertainty ($\sigma^*_{\bar{\alpha}}$). Bold values indicate results with more than 2σ significance.

Drift SAT-GND [%/decade]	10–20 km $\bar{\alpha}$	10–20 km $1\sigma^*_{\bar{\alpha}}$	20–30 km $\bar{\alpha}$	20–30 km $1\sigma^*_{\bar{\alpha}}$	30–45 km $\bar{\alpha}$	30–45 km $1\sigma^*_{\bar{\alpha}}$	Remark
SAGE II	$[-3,-1]$	1–3.5	$[-2,0]$	0.5–1	$[-2,+2]$	1–3	very stable
SAGE III	$[-15,-2]$	5–15	$[-10,0]$	3–6	no results		record too short
HALOE	$[-3,+5]$	1.5–6	**[-7, -1]**	1–2	$[-5,0]$	1.5–6	significant 20–30 km
UARS MLS	no UARS MLS data		$[-1,+5]$	2–4	$[-2,+3]$	3–12	record short
Aura MLS	$[-4,0]$	0.8–1.5	$[0,+3]$	0.5–1	$[-1,+3]$	1–6	very stable
POAM II	no results		$[-15,+15]$	9–20	no results		record too short
POAM III	$[-4,+10]$	2–10	$[-8,-2]$	2–4	no results		record sparse
OSIRIS	$[-5,+1]$	1–4	$[+1,+3]$	0.8–1	**[+1, +8]**	1–2.5	significant 36–44 km, indications 25–34 km
SMR	no SMR data		$[-3,+5]$	1.5–3	$[-15,+3]$	3–10	indications > 35 km
GOMOS	$[-12,-3]$	2–20	$[-4,-1]$	1.5–2.5	$[-1,+3]$	1.5–4	indications 15–25 km
MIPAS (OR)	$[-1,+3]$	1–2	$[0,+3]$	1–2.5	$[-4,+1]$	2.5–5	stable
SCIAMACHY	$[0,+4]$	1–2.5	$[+1,+3]$	0.8–1.5	**[-9, +1]**	1–5	significant 32–42 km, indications < 30 km
ACE-FTS	$[-5,0]$	3–7	$[-4,+3]$	2.5–3.5	no results		record sparse
MAESTRO	$[-7,+10]$	3–12	$[0,+6]$	3–4	no results		record sparse

Table 5. Overview of the bias of satellite ozone profile records relative to ozonesonde and lidar, in the upper troposphere and stratosphere. We present the range of the median relative difference (bias b) in each altitude bin, and whether there are any dependences on latitude and season that depart from the general tendency.

Bias SAT-GND [%]	<TP	TP–20 km	20–30 km	30–45 km	Remark
SAGE II	< -10	± 4	$[-5,+3]$	$[-1,+3]$	
SAGE III	< -15	± 3	± 3	$[0,+7]$	
HALOE	< -20	$[-15,0]$	$[-5,+2]$	$[0,+7]$	
UARS MLS	–	$[0,+10]$	$[0,+5]$	$[-5,+4]$	
Aura MLS	± 15	± 5	± 3	± 3	vertical oscillations in UTLS
POAM II	–	$[-7,-3]$	$[-10,-2]$	$[0,+10]$	
POAM III	–	$[-3,+10]$	$[-5,0]$	$[0,+5]$	
OSIRIS	$[-15,-10]$	$[-10,0]$	± 4	$[0,+5]$	$+5$ % at 22 km
SMR	–	$[0,+10]$	$[-7,-3]$	$[-5,-10]$	negative bias above TP+10 km
GOMOS	-20 (N) $+20$ (S)	$[-10,0]$ (N) $[+5,+15]$ (S)	± 3	± 3	N/S sign change troposphere larger bias in Arctic
MIPAS OR	$> +20$	$[0,+5]$	$[0,+8]$	$[+3,+9]$	persistent positive bias
SCIAMACHY	$> +20$	± 7	$[-5,+15]$	$[0,+15]$	(a) Arctic seasonality, (b) large + bias towards S hemisphere
ACE-FTS	$> +10$	± 3	± 3	$[0,+5]$	
MAESTRO	–	$[-20,0]$	$[-5,+3]$	± 4	sharp change < 15 km

Table 6. Decadal stability of merged ozone profile records (Level-3) estimated from our ground-based assessment of the stability of the contributing Level-2 records, in two time periods and three layers of the stratosphere. These drift values could serve as 1σ systematic uncertainty in trend studies. Nevertheless, in expectation of more rigourous analyses, the estimates below should be considered with care, as they may overestimate the actual drift.

Stability merged record [%/decade]	Pre-1997			Post-1998		
	LS	MS	US	LS	MS	US
SAGE II–GOMOS	1–1.5	1	1–1.5	3–5	1.5–2	1.5–2
SAGE II–OSIRIS	1–1.5	1	1–1.5	2	2	3–5
GOZCARDS	1–1.5	1–1.5	1–1.5	1.5–2	1.5–2	1.5–2
SWOOSH	1–1.5	1–1.5	1–1.5	1.5–2	1.5–2	1.5–2
Ozone_cci	no data			2–2.5	2	2–3.5

*LS: 10–20 km ($\sim$ 50–250 hPa), MS: 20–30 km ($\sim$ 10–50 hPa), US: 30–45 km ($\sim$ 1–10 hPa).

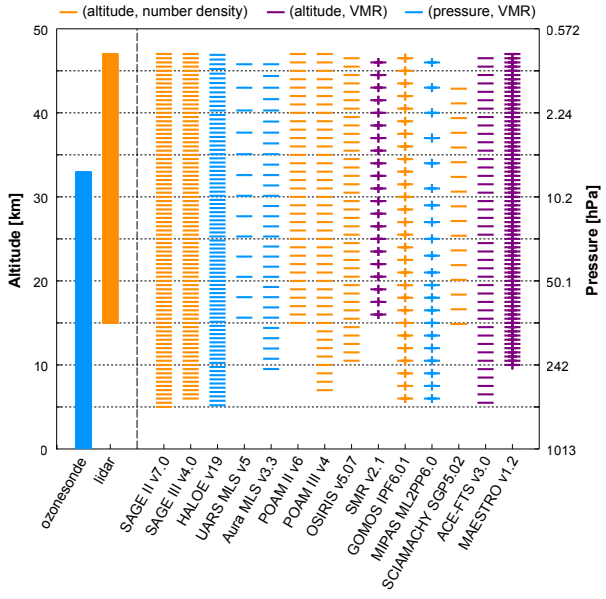

Figure 1. Overview of the native representation of the ozone profile records (see legend). A vertical band defines the approximate range for the ground-based data sets, while individual levels are shown for satellite profiles. Only the levels considered in our analyses are depicted. Profile-dependent vertical grids are marked with small vertical bars. Differences between geometric and geopotential height are neglected.

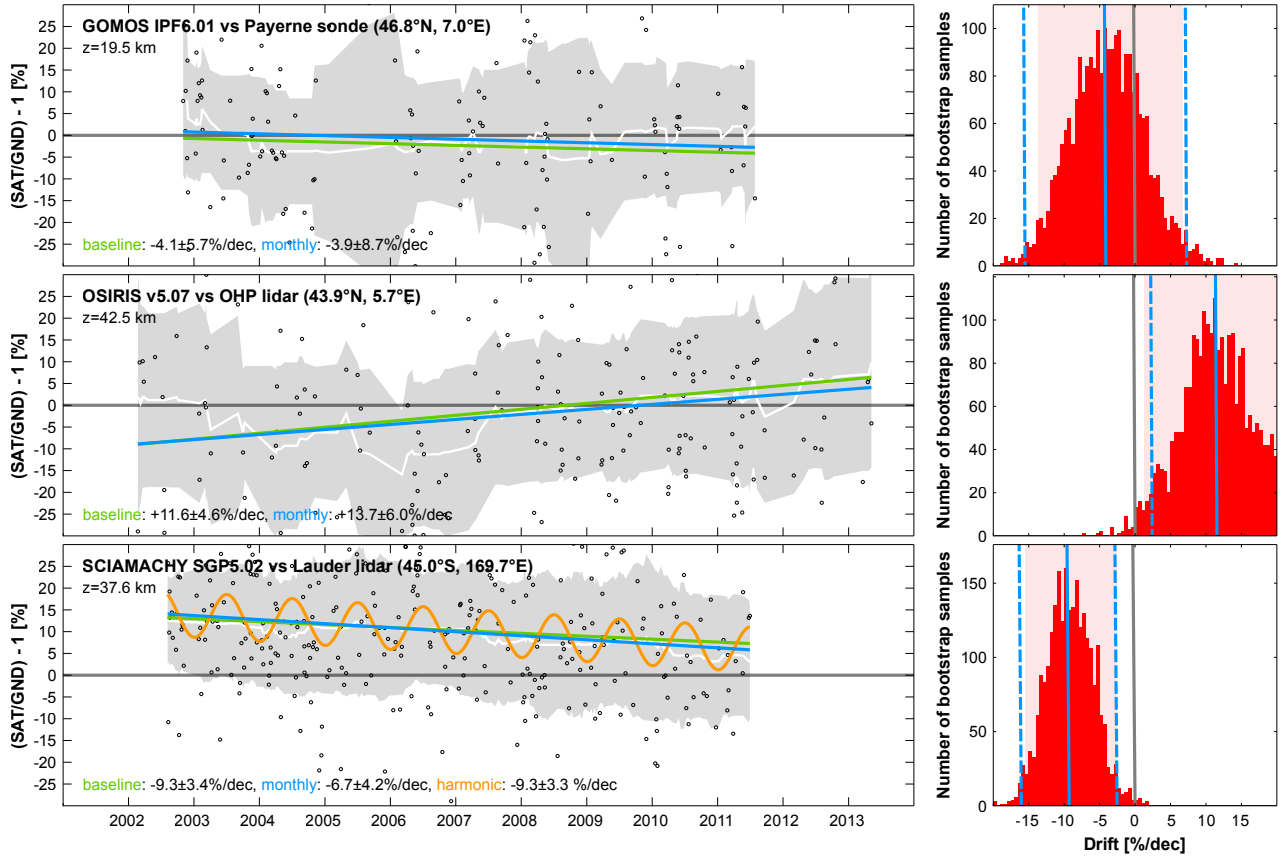

Figure 2. (Left) Time series of the ozone comparisons for GOMOS versus Payerne ozonesonde at 19.5 km (top), for OSIRIS versus OHP lidar at 42.5 km (centre), and for SCIAMACHY versus Lauder lidar at 37.6 km (bottom). A one-year running median filter is applied to highlight long-term dependence (white line and 1σ shaded area). The blue line depicts the baseline regression model, while the green and orange lines show cross-checks (see text). The estimated drift $\hat{\alpha}$ and its $1\hat{\sigma}_\alpha$ uncertainty is mentioned at the bottom of each panel. (Right) Distribution of the drift obtained from 2500 bootstrapped samples of the time series on the left. The light red zone marks the 95 % interpercentile of the drift distribution, which should be compared to the analytical derivation $\hat{\alpha} \pm 2\hat{\sigma}_\alpha$ (vertical blue lines).

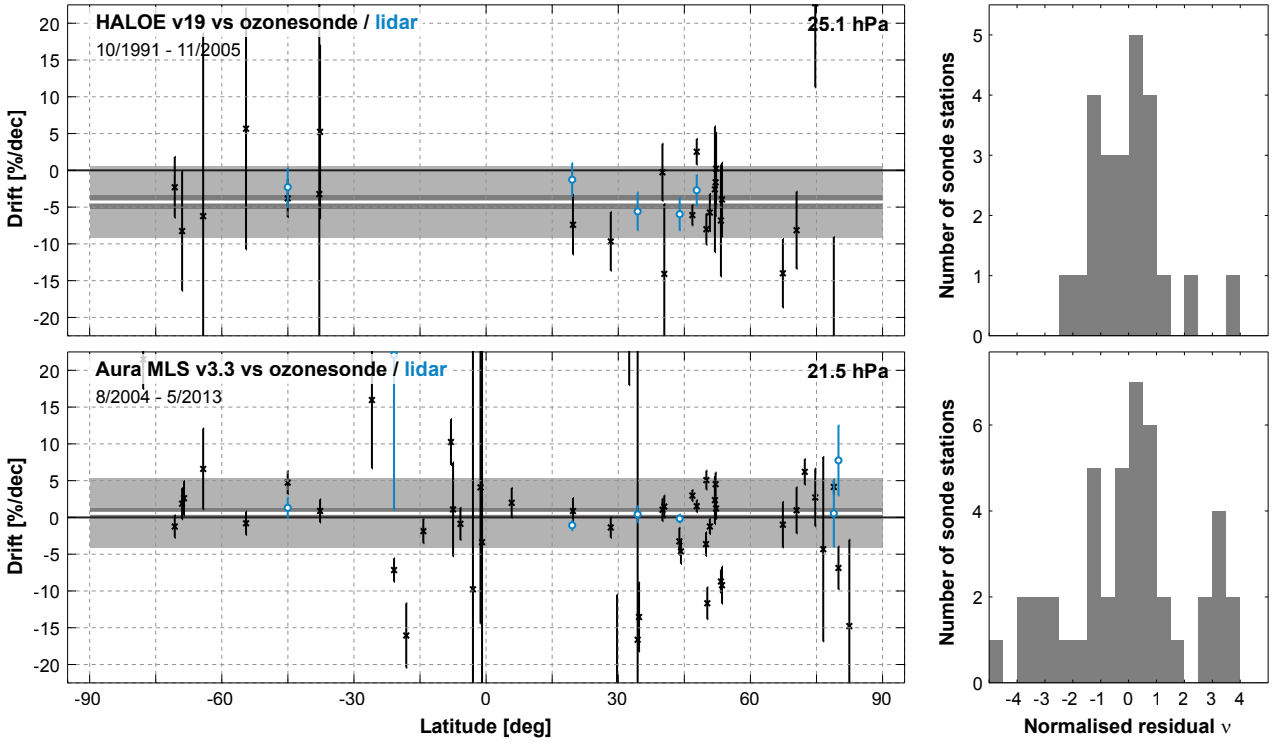

Figure 3. (Left) Drift $\hat{\alpha}_j$ and one sigma uncertainty $\hat{\sigma}_{\alpha,j}$ of HALOE (top) and Aura MLS (bottom) around 25 km relative to co-located observations at each station in the NDACC/GAW/SHADOZ ozonesonde (black) and NDACC lidar (blue) networks. Also shown are the weighted mean for the ozonesonde network $\bar{\alpha}$ (horizontal white line), its adjusted uncertainty $\sigma^*_{\bar{\alpha}}$ (dark grey area) and the standard deviation of the ensemble of single sonde site drift estimates (light grey area). (Right) Corresponding distribution of normalised residuals ν from the ozonesonde network.

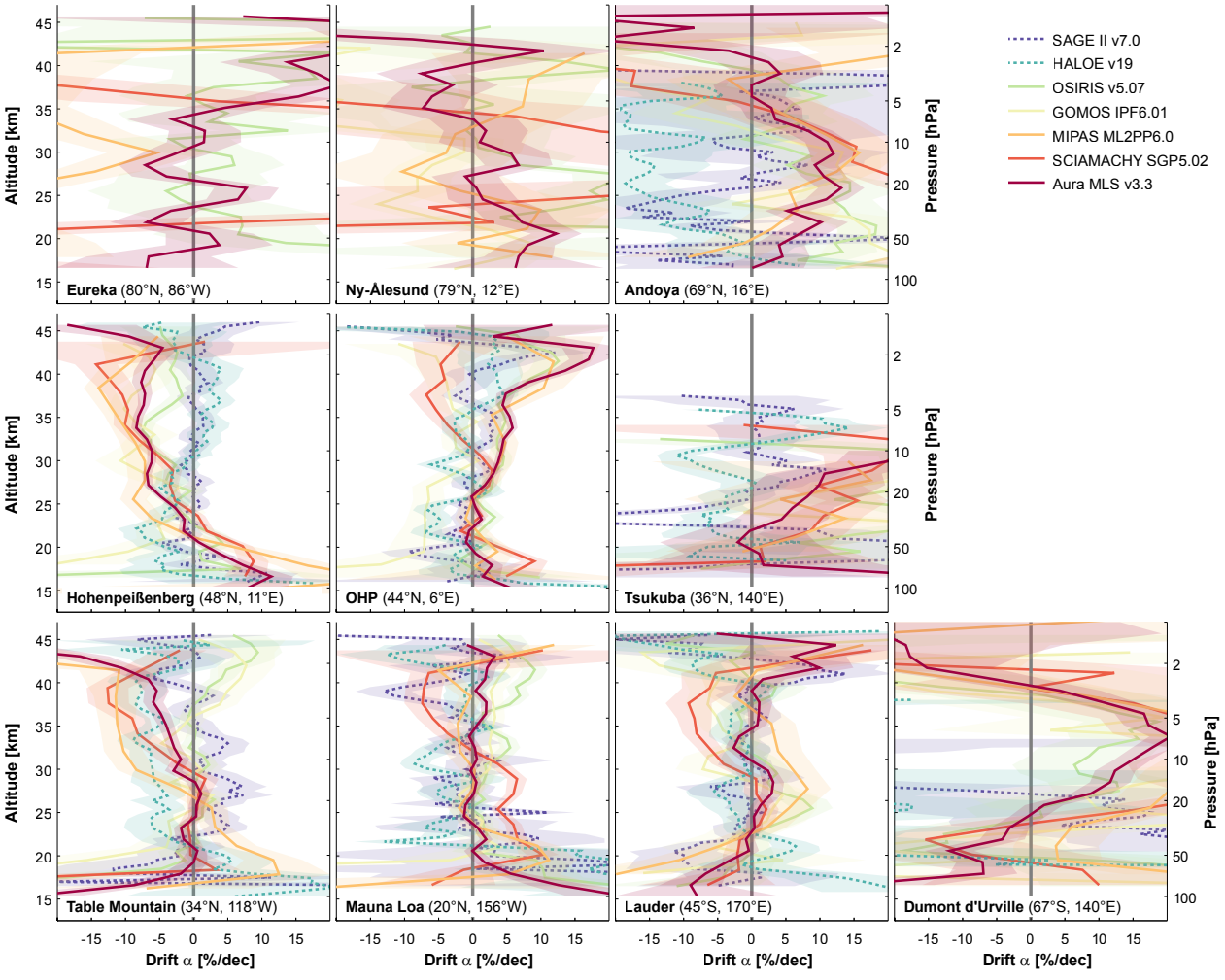

Figure 4. Comparison of the vertical structure of the drift α of two historic (dashed lines) and five recent (solid) satellite records relative to stratospheric lidar observations at ten NDACC stations. The shaded area represents the unadjusted 68% confidence interval, which does not include possible uncertainties from differences in sampling or from inhomogeneities in the lidar network. The analysis is performed in the native profile representation of each satellite record.

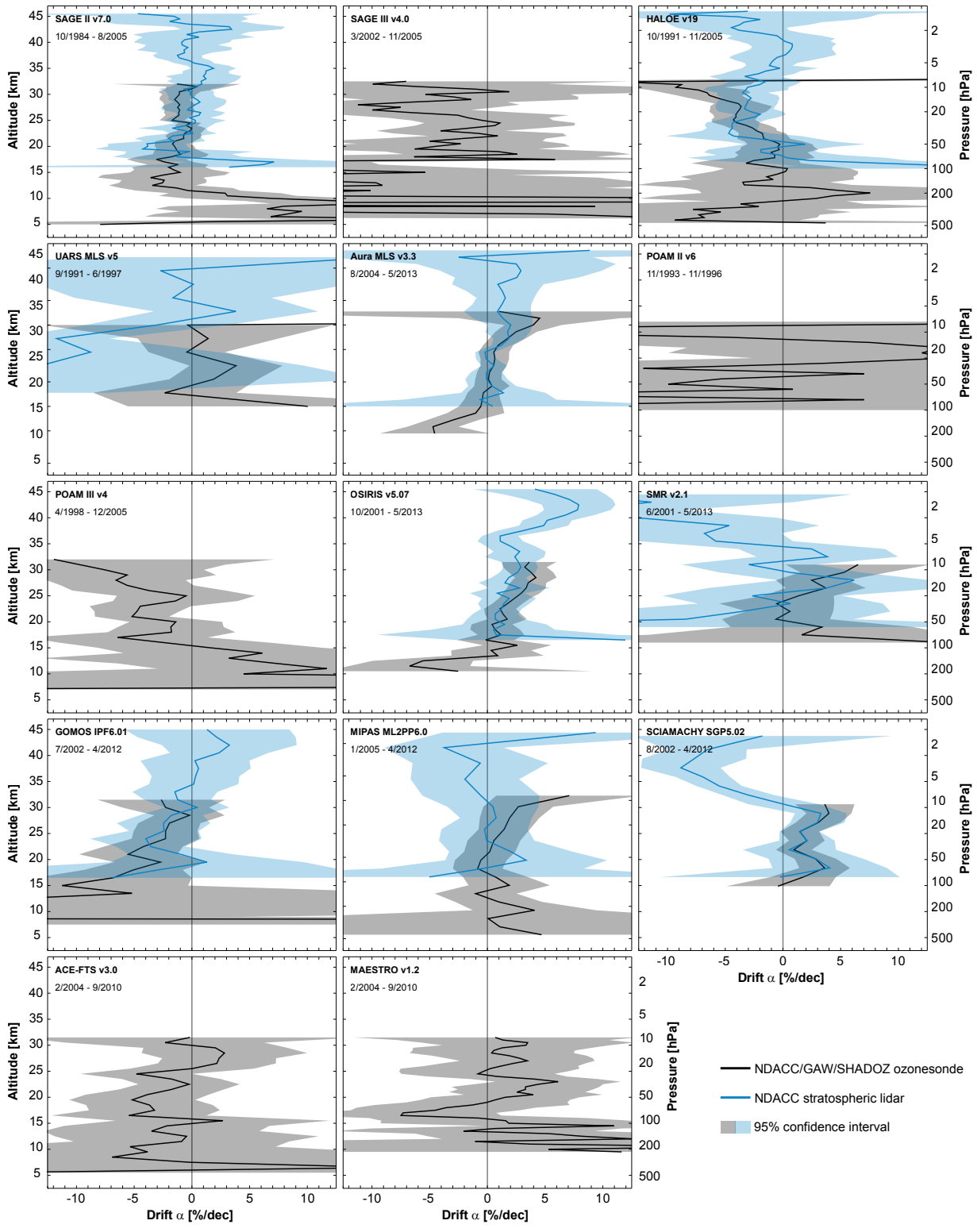

Figure 5. Average drift $\bar{\alpha}$ of each satellite record relative to the entire ozonesonde (black) and lidar network (blue). The shaded region indicates the χ^2-adjusted 95 % confidence interval for the drift. The analysis is performed in the native profile representation of each satellite record. Y-axis tick marks correspond to the native vertical coordinate of the satellite record and may be offset from the axis labels in case the coordinate differs.

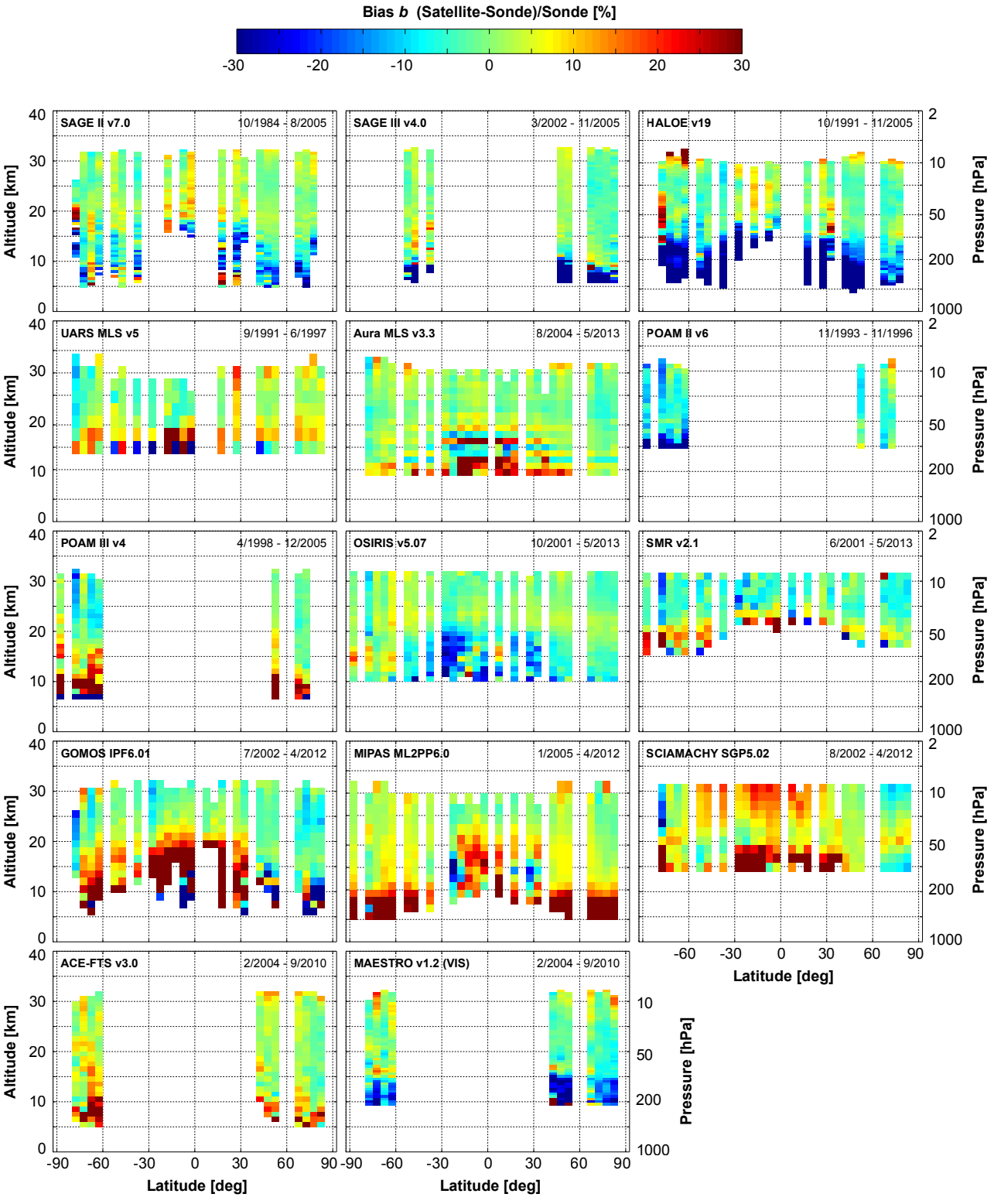

Figure 6. Meridional structure of the bias b relative to the ozonesonde network for all fourteen satellite ozone profile records. The analysis is performed in the native profile representation of each satellite record. Only bins with more than 5 comparison pairs are shown.

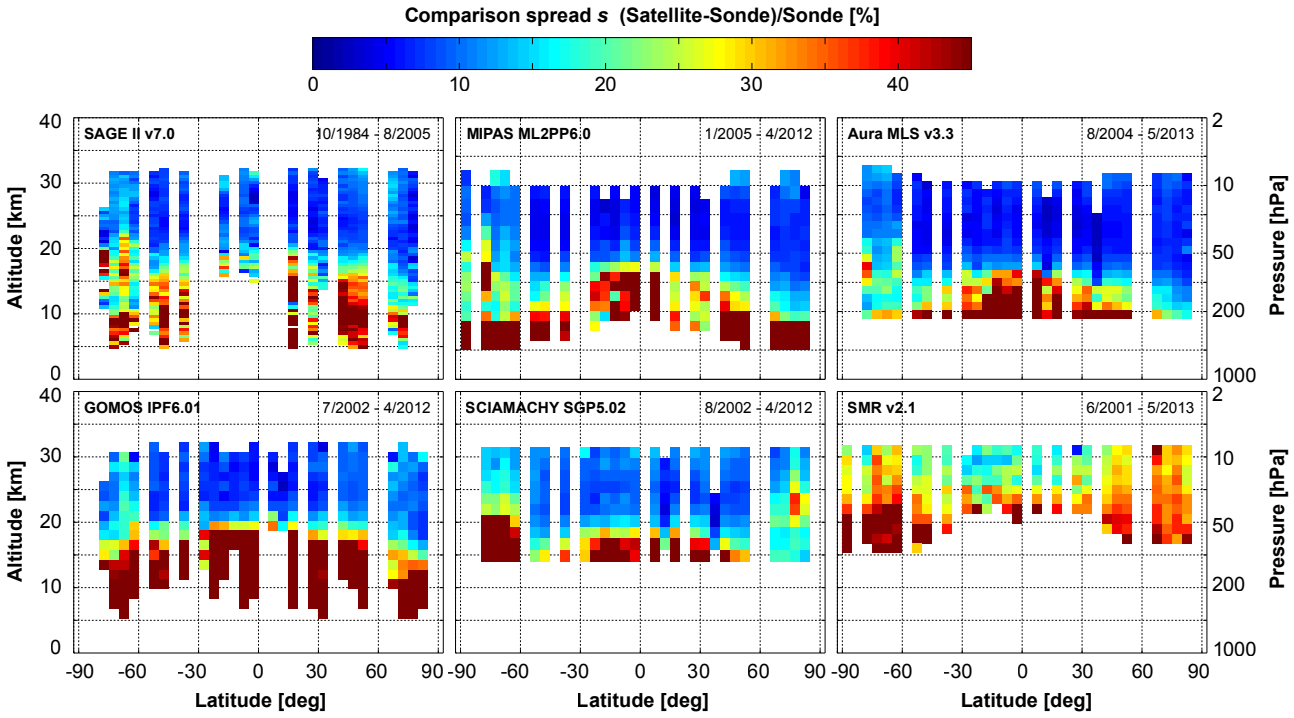

Figure 7. Meridional structure of the comparison spread s of six satellite records relative to the ozonesonde network. The analysis is performed in the native profile representation of each satellite record. Only bins with more than 5 comparison pairs are shown.

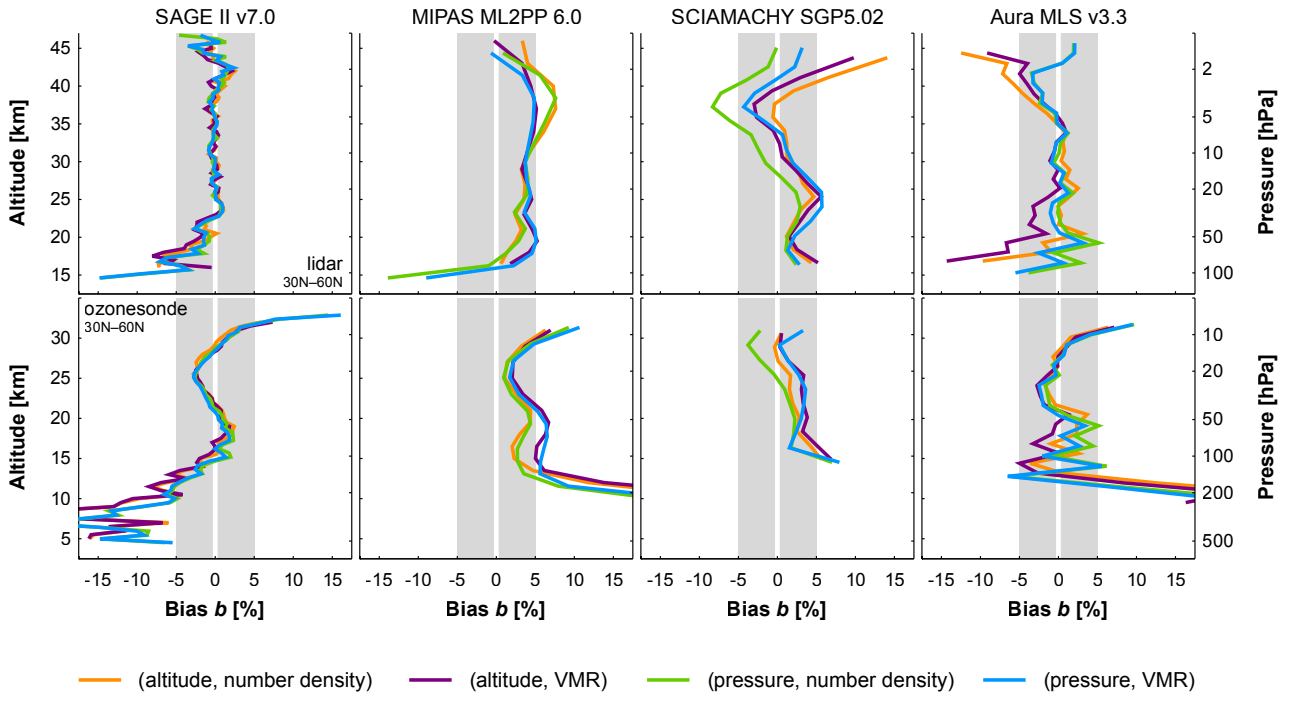

Figure 8. Dependence on profile representation of the bias of four satellite data records relative to lidar (top row) and ozonesonde data (bottom) at northern mid-latitudes. The satellite profiles are converted using the auxiliary information provided in the respective data files. The Supplement contains the results for all satellite records in five latitude bands.

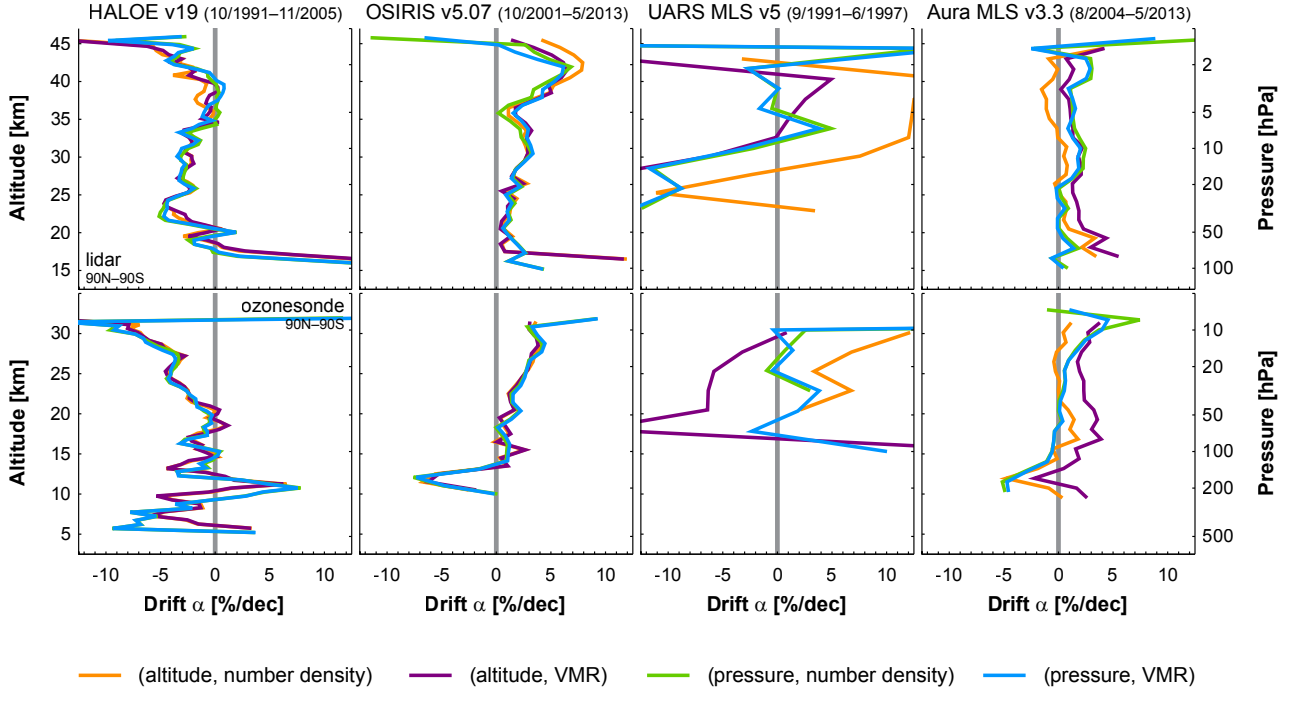

Figure 9. Dependence on profile representation of the network-averaged drift $\bar{\alpha}$ of four satellite data records from comparisons to lidar (top row) and ozonesonde (bottom). The satellite profiles are converted using the auxiliary information provided in the respective data files. The Supplement contains the results for all satellite records.

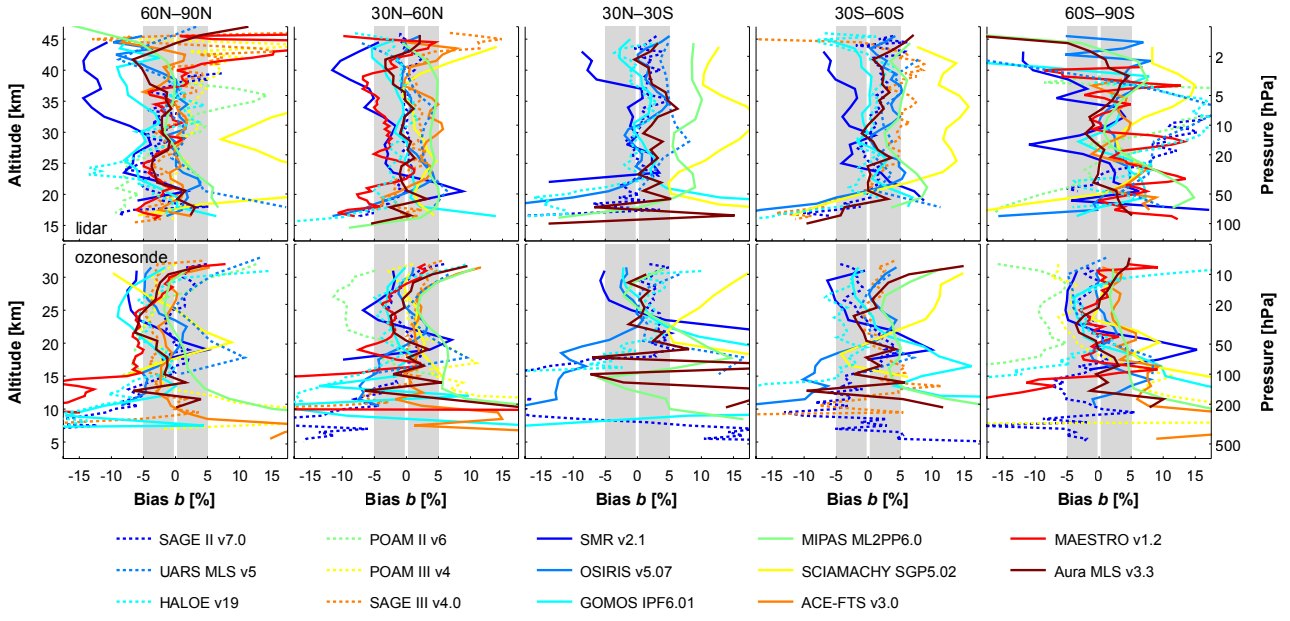

Figure 10. Overview of the bias b of all satellite ozone records relative to the stratospheric ozone lidar (top row) and ozonesonde (bottom) network in five latitude bands (columns). Dashed lines indicate instruments that ceased operations prior to 2006. The analysis is done in the native profile representation of each satellite record. Most satellite records agree within ± 5 % with the ground-based data in the middle and upper stratosphere (grey shaded area).

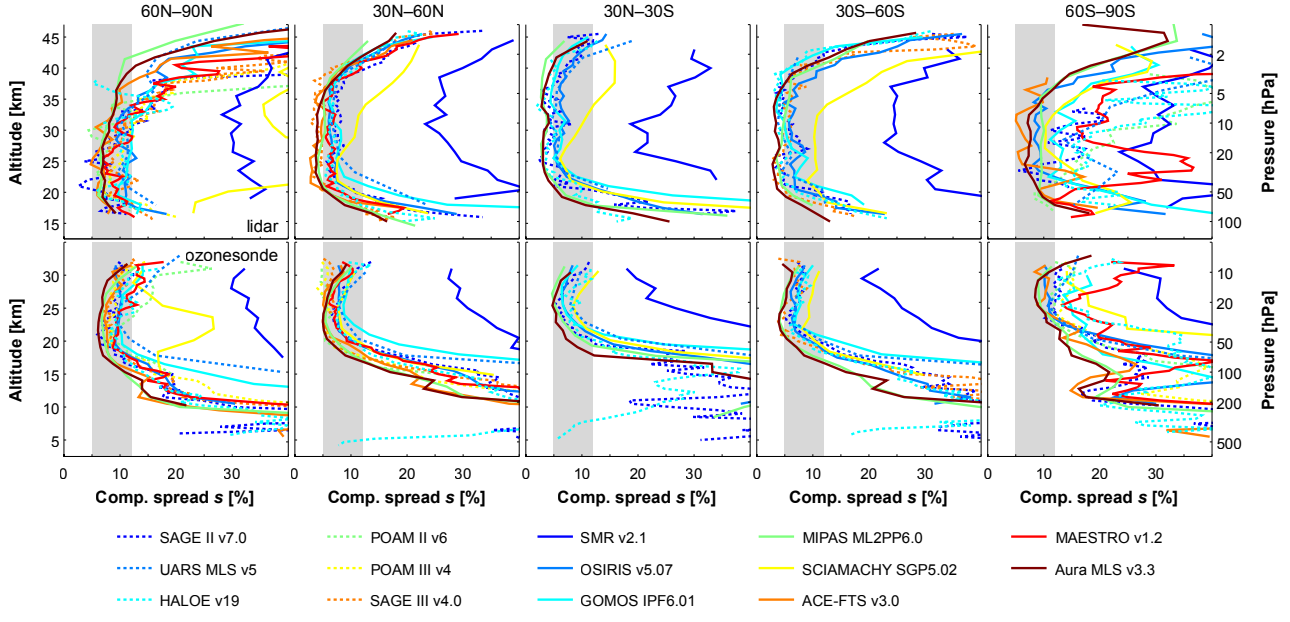

Figure 11. Similar as Fig. 10, but for the comparison spread s. The spread is mostly between 5–12 % (grey shaded area) in the middle and upper stratosphere.

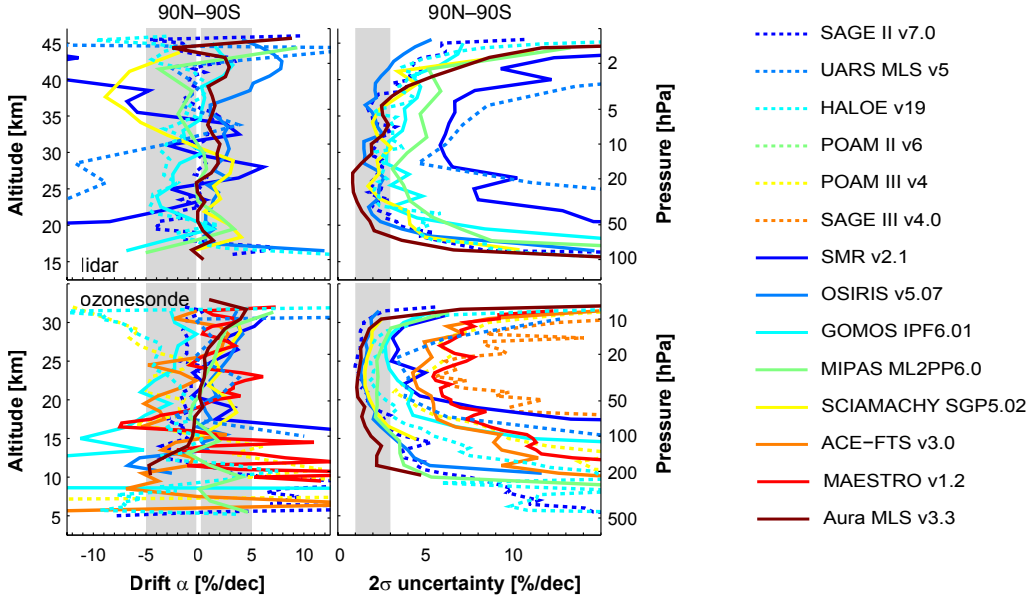

Figure 12. Overview of the drift $\bar{\alpha}$ (left column) and its 2σ uncertainty (right) of all satellite ozone records relative to the lidar (top row) and ozonesonde (bottom) networks. Dashed lines indicate instruments that ceased operations prior to 2006. The analysis is done in the native system of each satellite record. The decadal stability of most records remains within $\pm 5\,\%\,\mathrm{decade}^{-1}$ in the middle and upper stratosphere (shaded area). POAM II and SAGE III drift results are not displayed in the lower left panel to avoid clutter.

1 Introduction

This document provides supplementary figures which could not be shown in the main manuscript due to space limitations.

- Fig. S1: latitude-time cross section of the co-location samples used in the analysis.
- Fig. S2: vertical profile of adjustment factor κ for ozonesonde and lidar networks, and all satellite records.
- Fig. S3–S16: bias, comparison spread and network-averaged drift results in four ozone profile representations. One figure per satellite record; results are split up per type of correlative instrument and latitude band to allow a more quantitative reading of the results.
- Fig. S17: sensitivity of the network-averaged drift results to several regression analysis parameters.
- Fig. S18: sensitivity of the network-averaged drift results to selection of ground sites.
- Fig. S19: bias relative to ozonesonde and lidar of sunrise and sunset observations by solar occultation instruments.
- Fig. S20: month-dependence of the bias and comparison spread of seven sounders relative to Arctic ozonesondes.

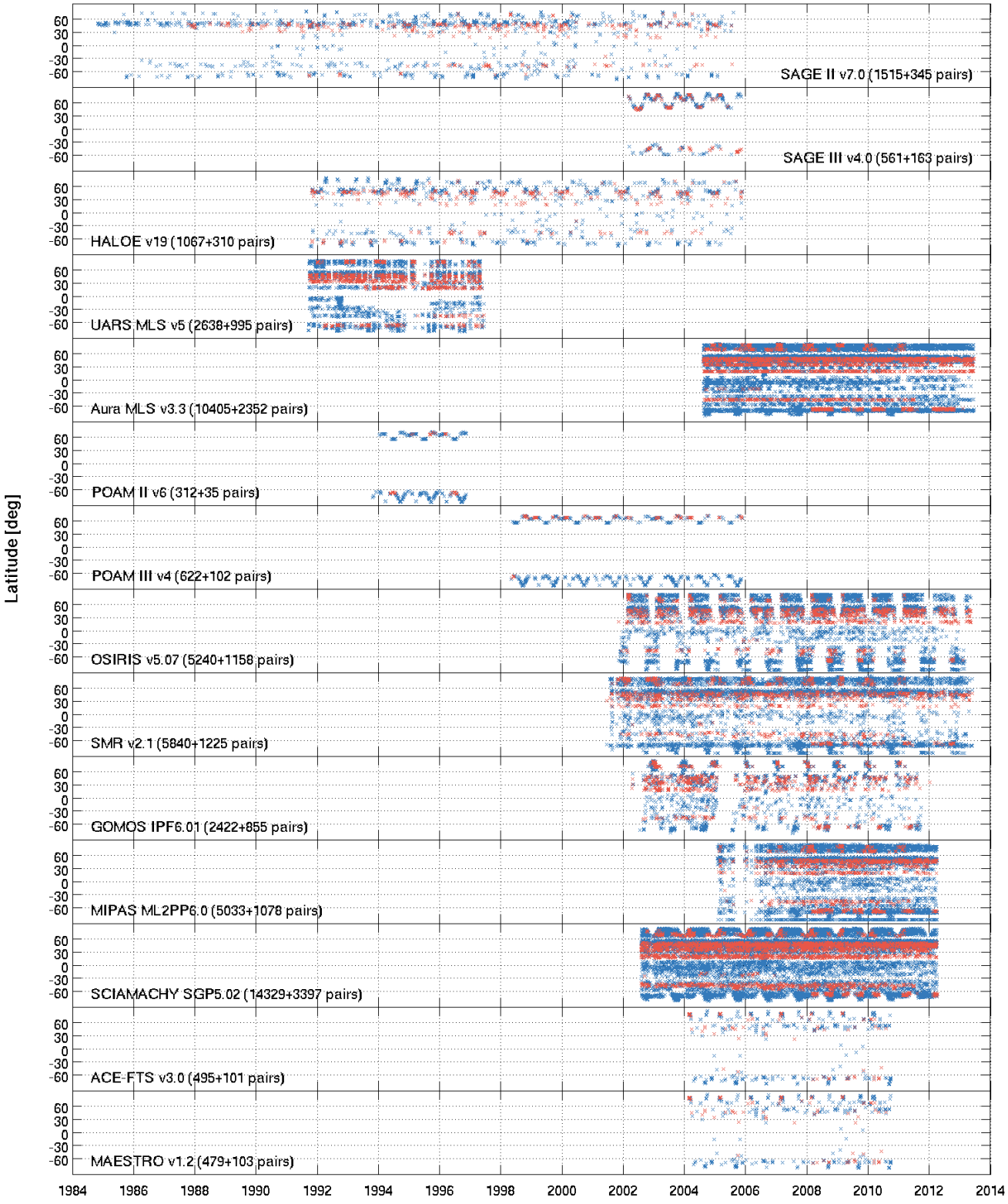

Figure S1. Latitude-time cross section of the co-locations of NDACC/GAW/SHADOZ ozonesonde (blue) and NDACC stratospheric ozone lidar (red) with fourteen limb/occultation ozone profile records. Latitude and time correspond to that of satellite profile. Total number of co-locating pairs for sonde and lidar are mentioned separately between brackets.

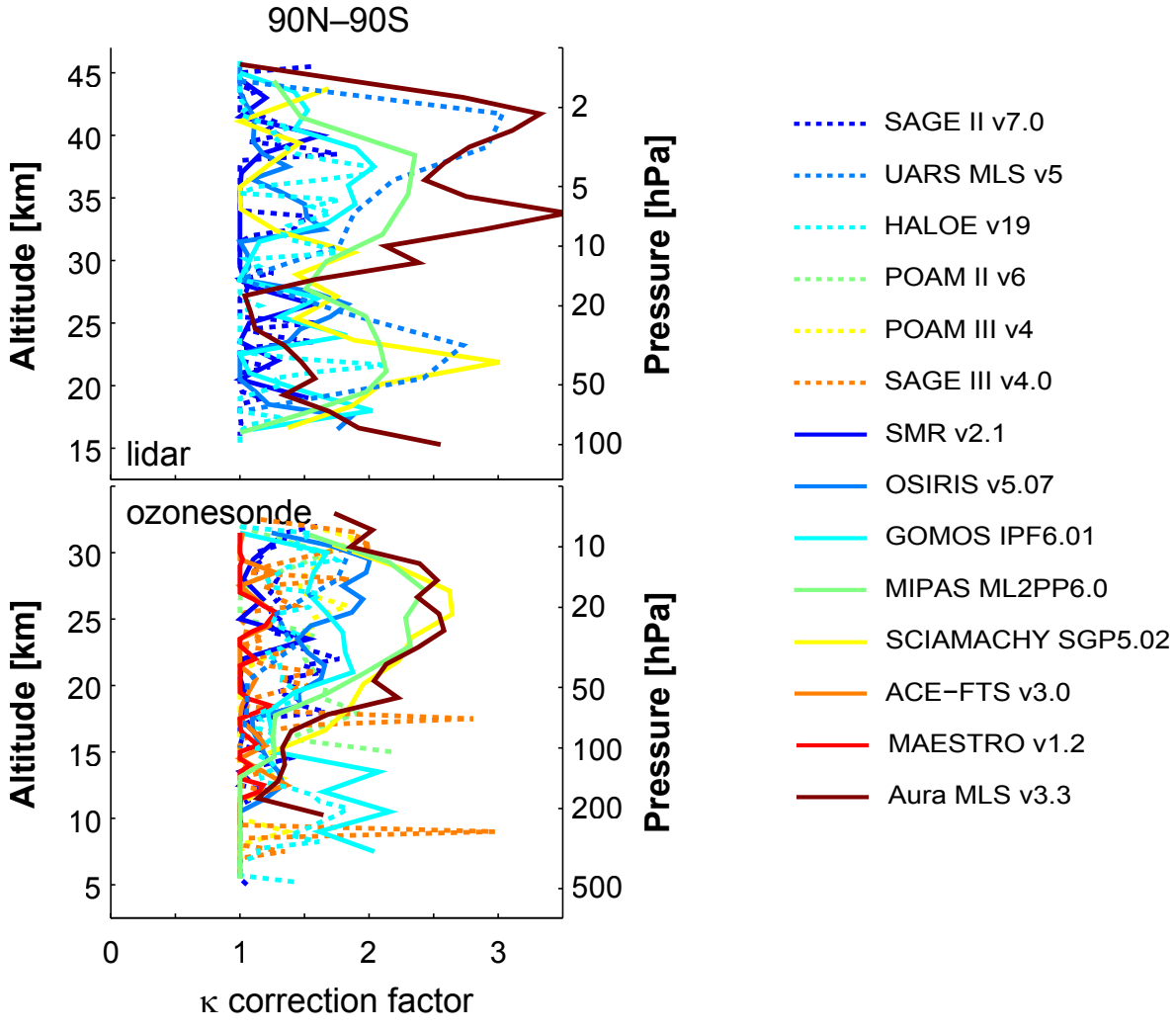

Figure S2. Adjustment factor κ to incorporate unaccounted sources of uncertainty (sampling, ground-network homogeneity, ...) in the estimation of network-averaged drift of limb/occultation ozone profile data relative to lidar (top) and ozonesonde (bottom) measurements. See Sect. 4.1.2 for more details.

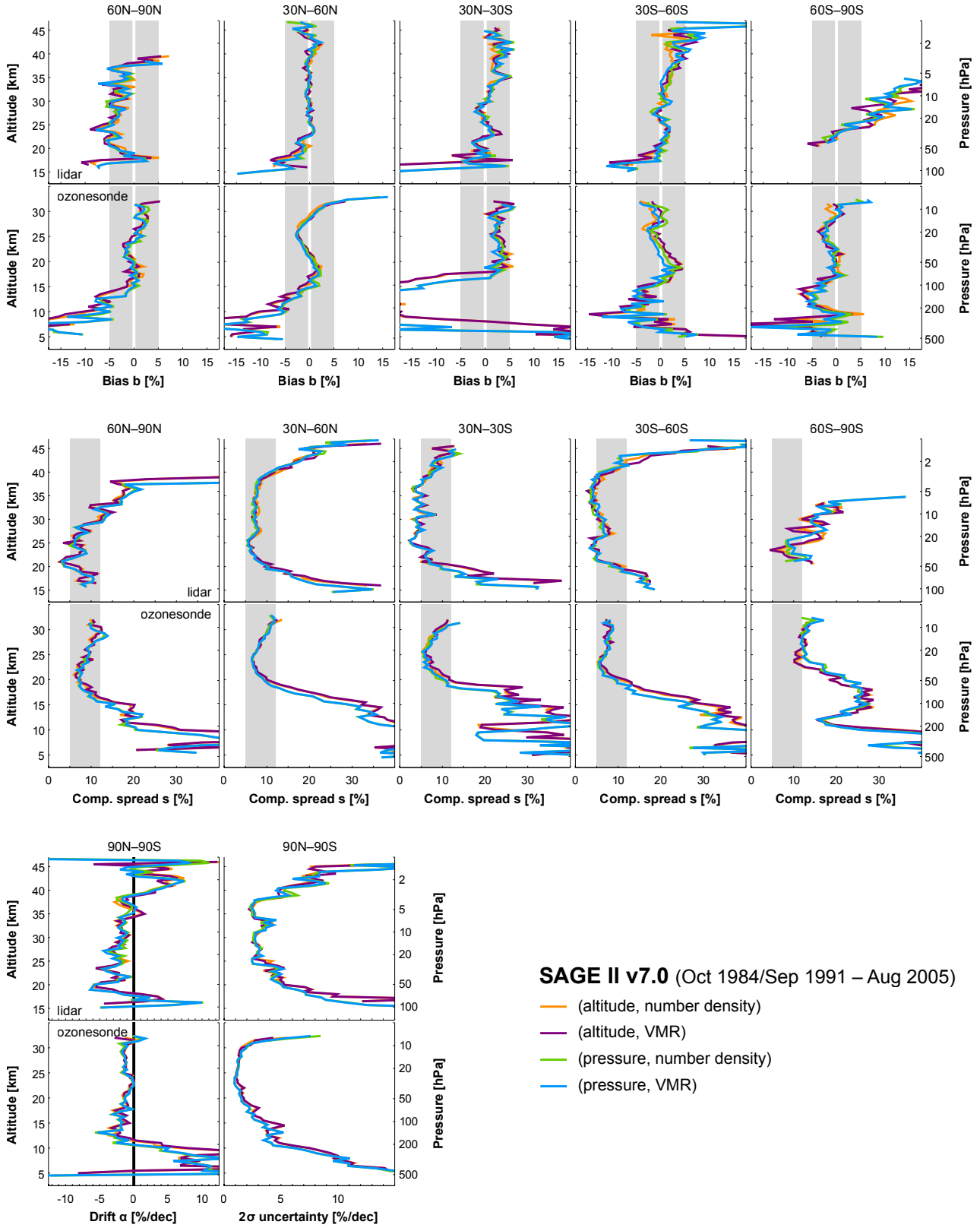

Figure S3. SAGE II v7.0 ozone data quality in different profile representations (see legend): bias (top part), comparison spread (middle) and drift (bottom) from comparisons to NDACC/GAW/SHADOZ ozonesonde and NDACC stratospheric lidar instruments. Only latitude bands with at least 10 comparison pairs are shown. The nominal representation of SAGE II is ozone number density on fixed altitude levels. Satellite and ozonesonde profiles are converted using auxiliary information included in the ozone data files, lidar profiles are converted using ERA-Interim fields. More information in Sect. 6. The SAGE II versus lidar results shown here cover 1991–2005, since we had no ERA-Interim data at our disposal to cover the earlier years. The SAGE II versus ozonesonde results cover 1984–2005.

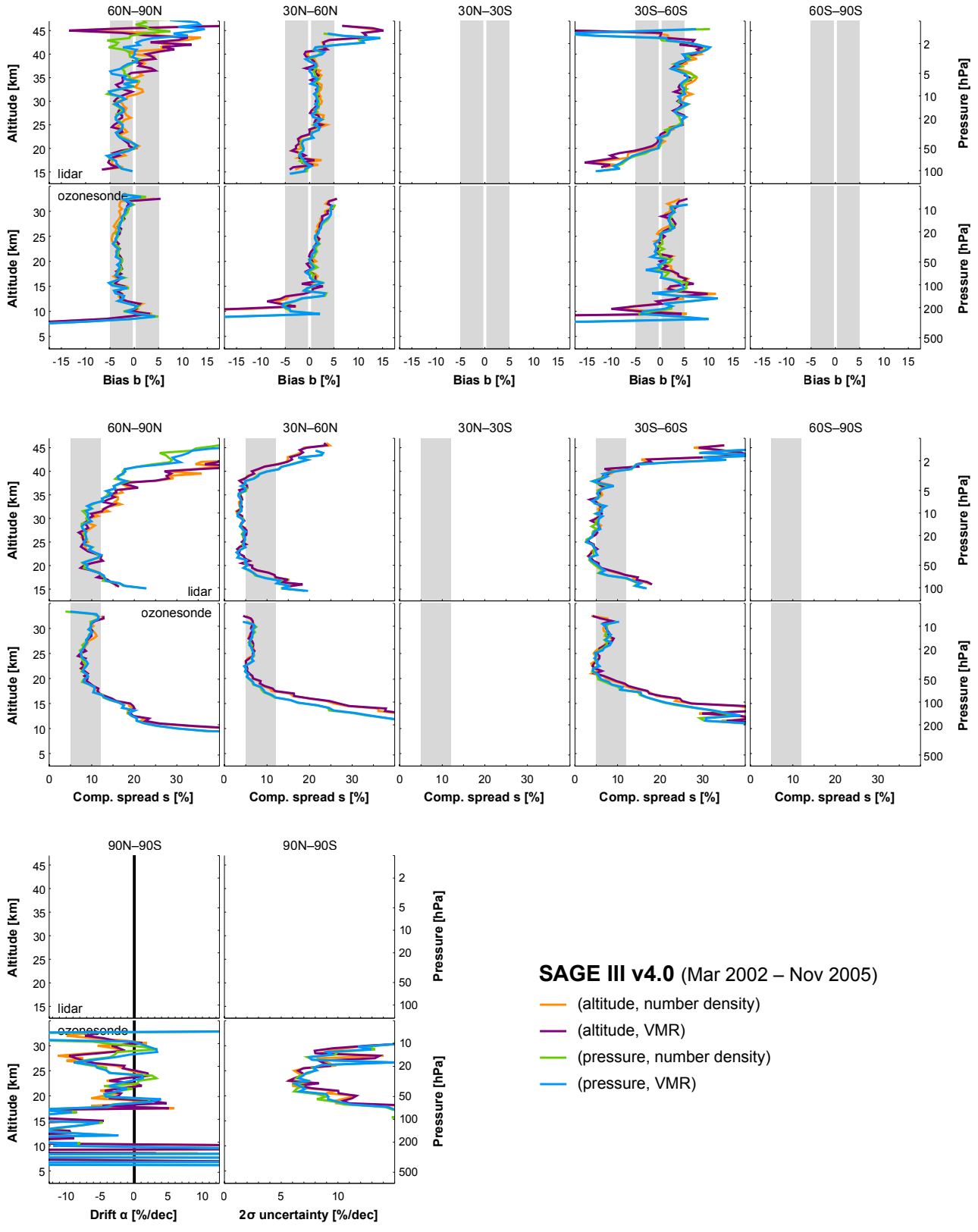

Figure S4. Same as Fig. S3, but for SAGE III v4.0 (nominal representation: ozone number density on fixed altitude levels).

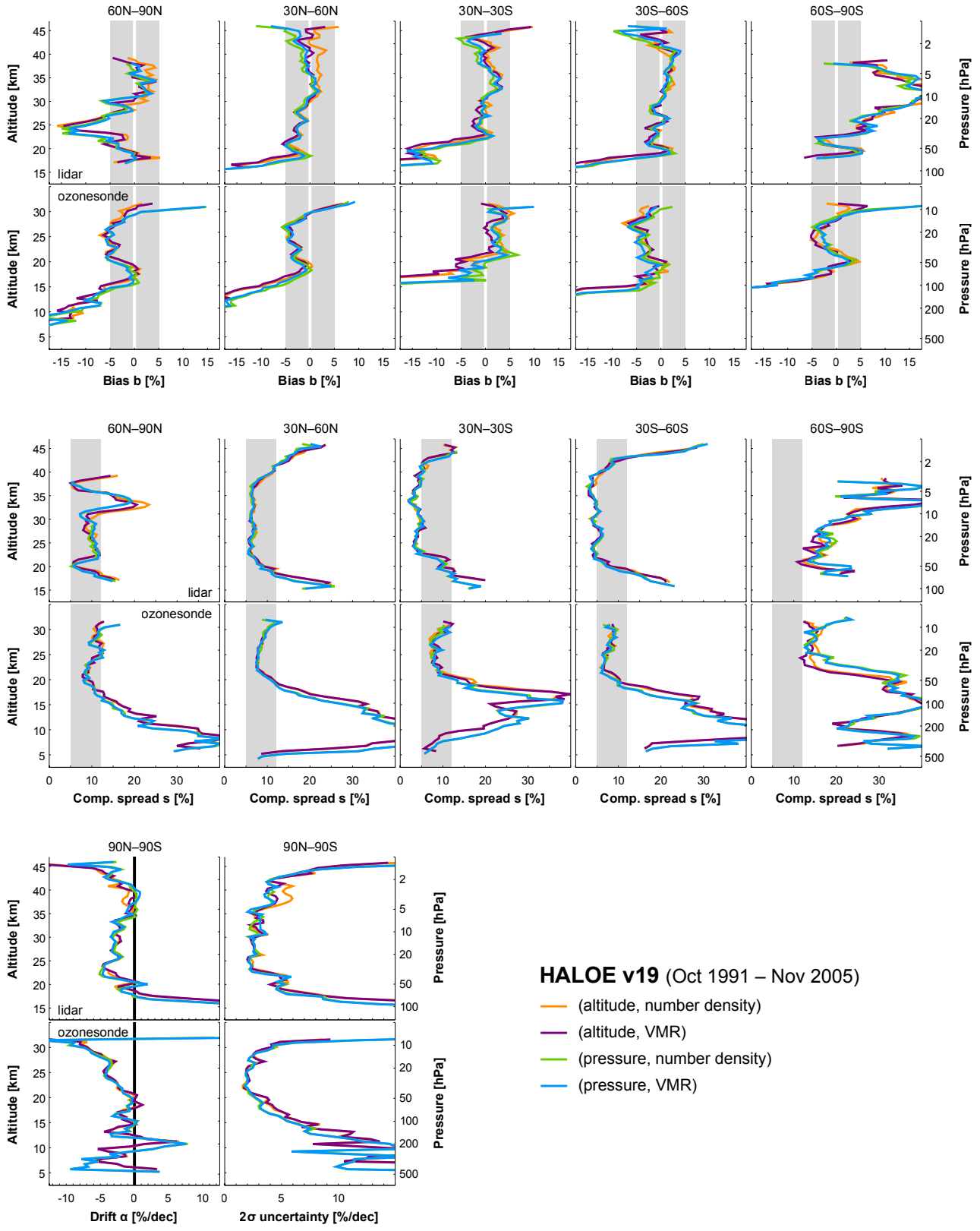

Figure S5. Same as Fig. S3, but for HALOE v19 (nominal representation: ozone VMR on fixed pressure levels).

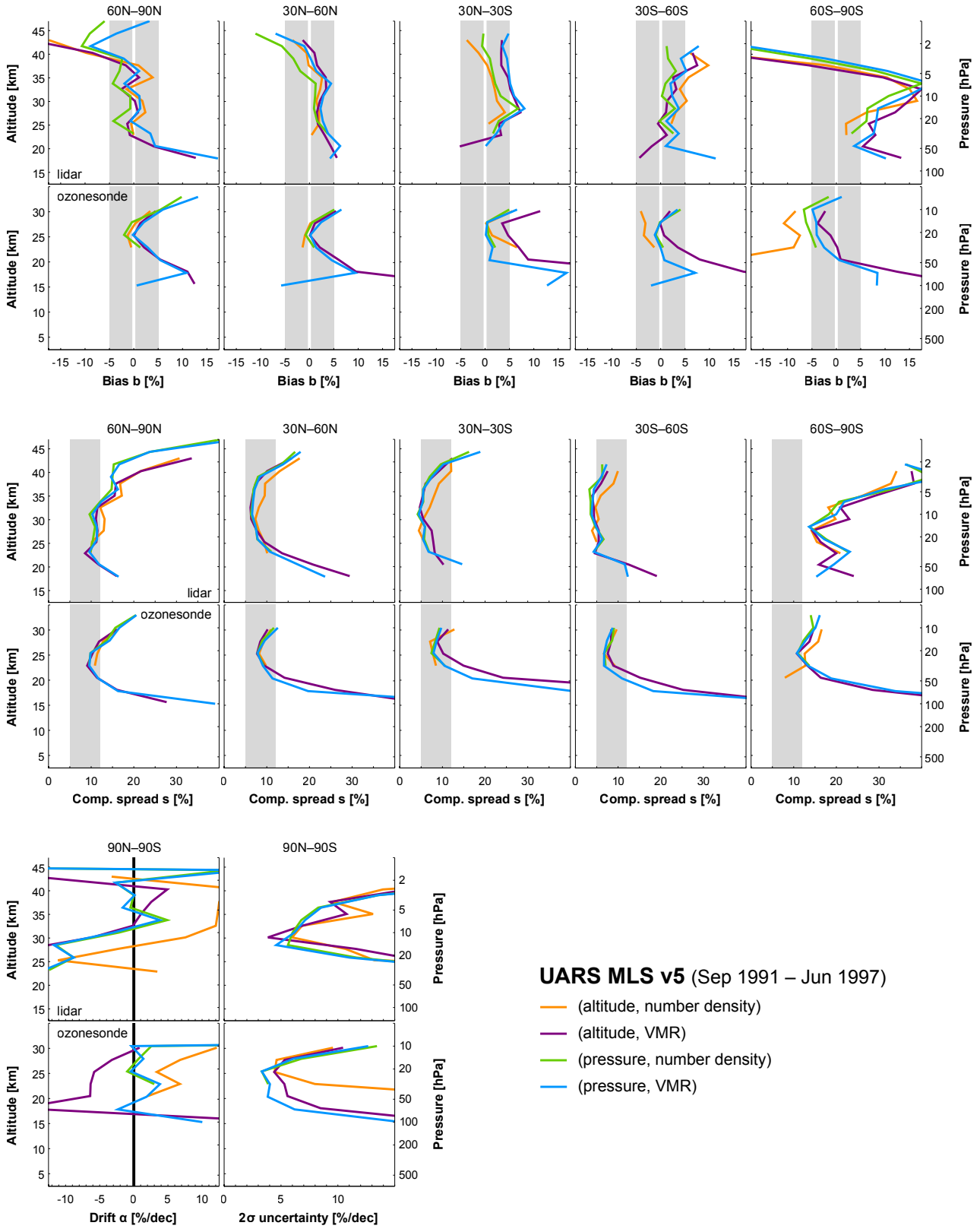

Figure S6. Same as Fig. S3, but for UARS MLS v5 (nominal representation: ozone VMR on fixed pressure levels).

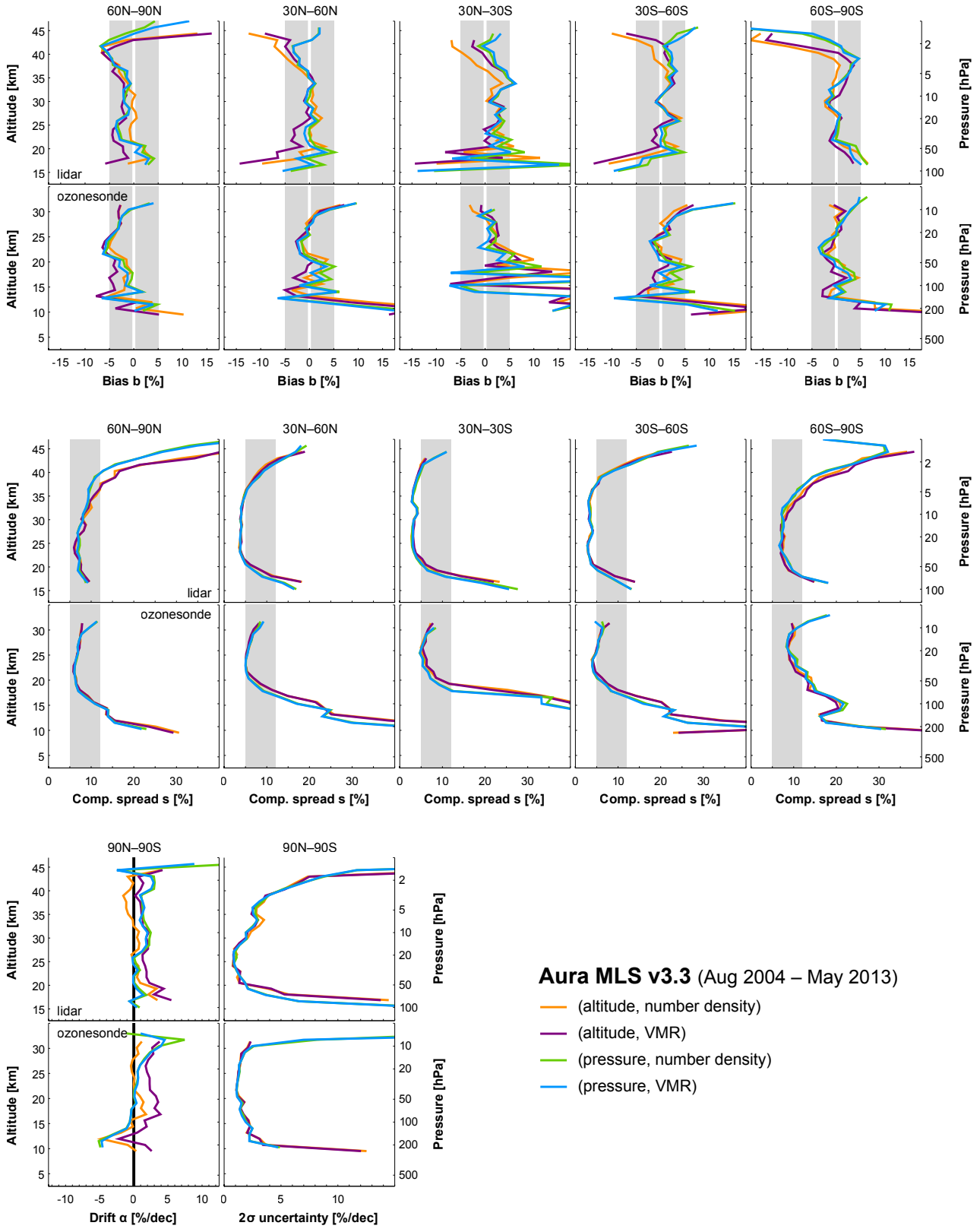

Figure S7. Same as Fig. S3, but for Aura MLS v3.3 (nominal representation: ozone VMR on fixed pressure levels).

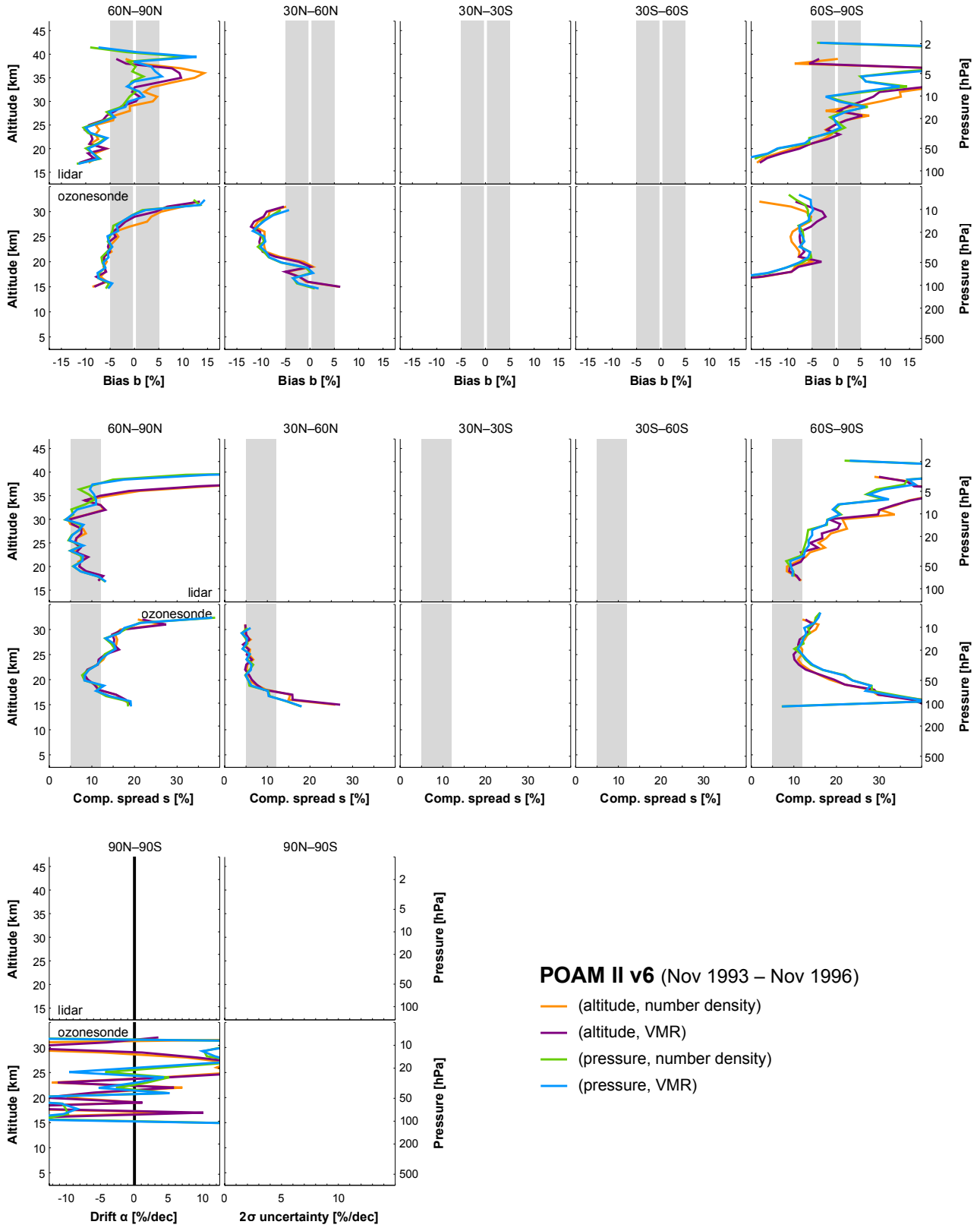

Figure S8. Same as Fig. S3, but for POAM II v6 (nominal representation: ozone number density on fixed altitude levels).

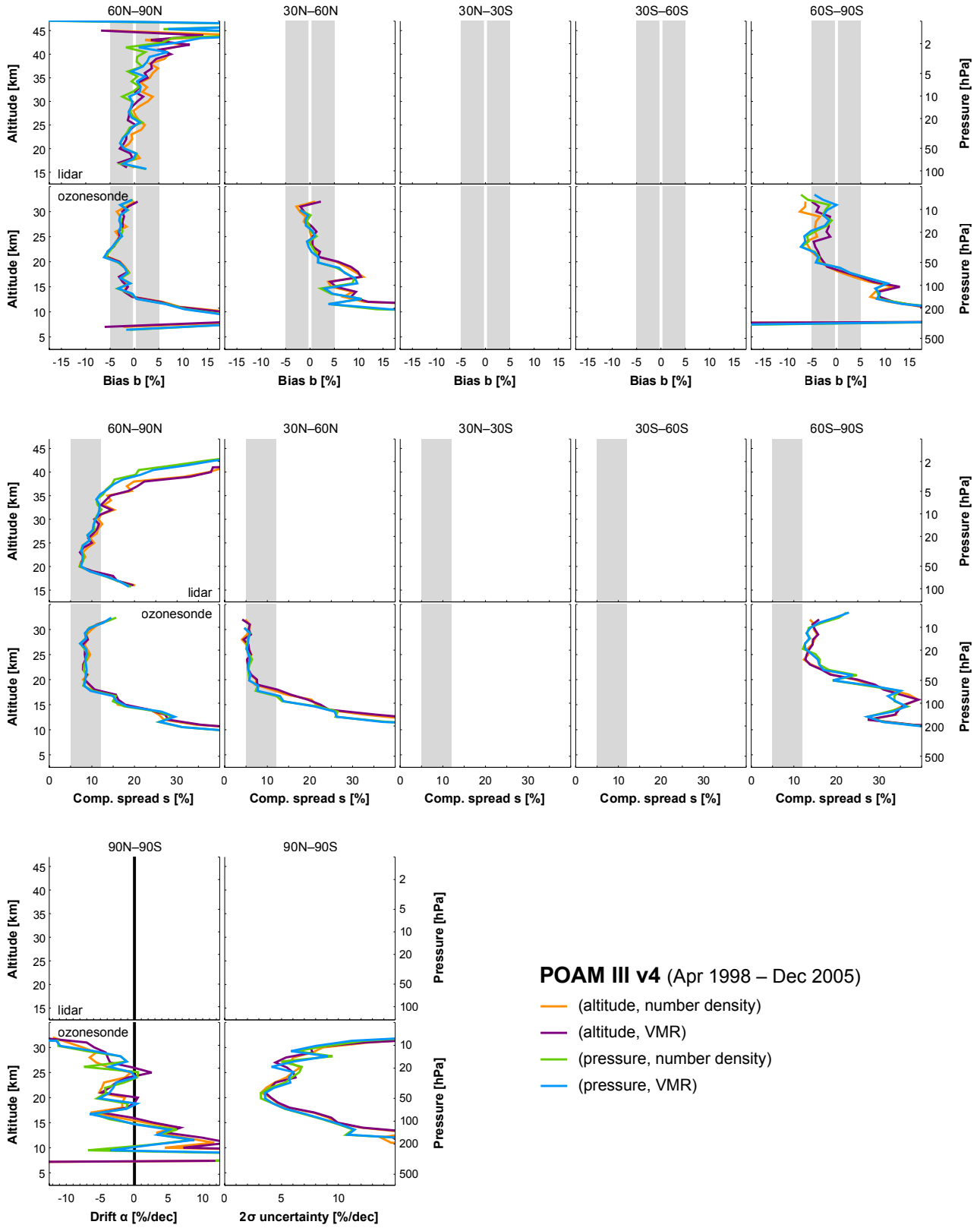

Figure S9. Same as Fig. S3, but for POAM III v4 (nominal representation: ozone number density on fixed altitude levels).

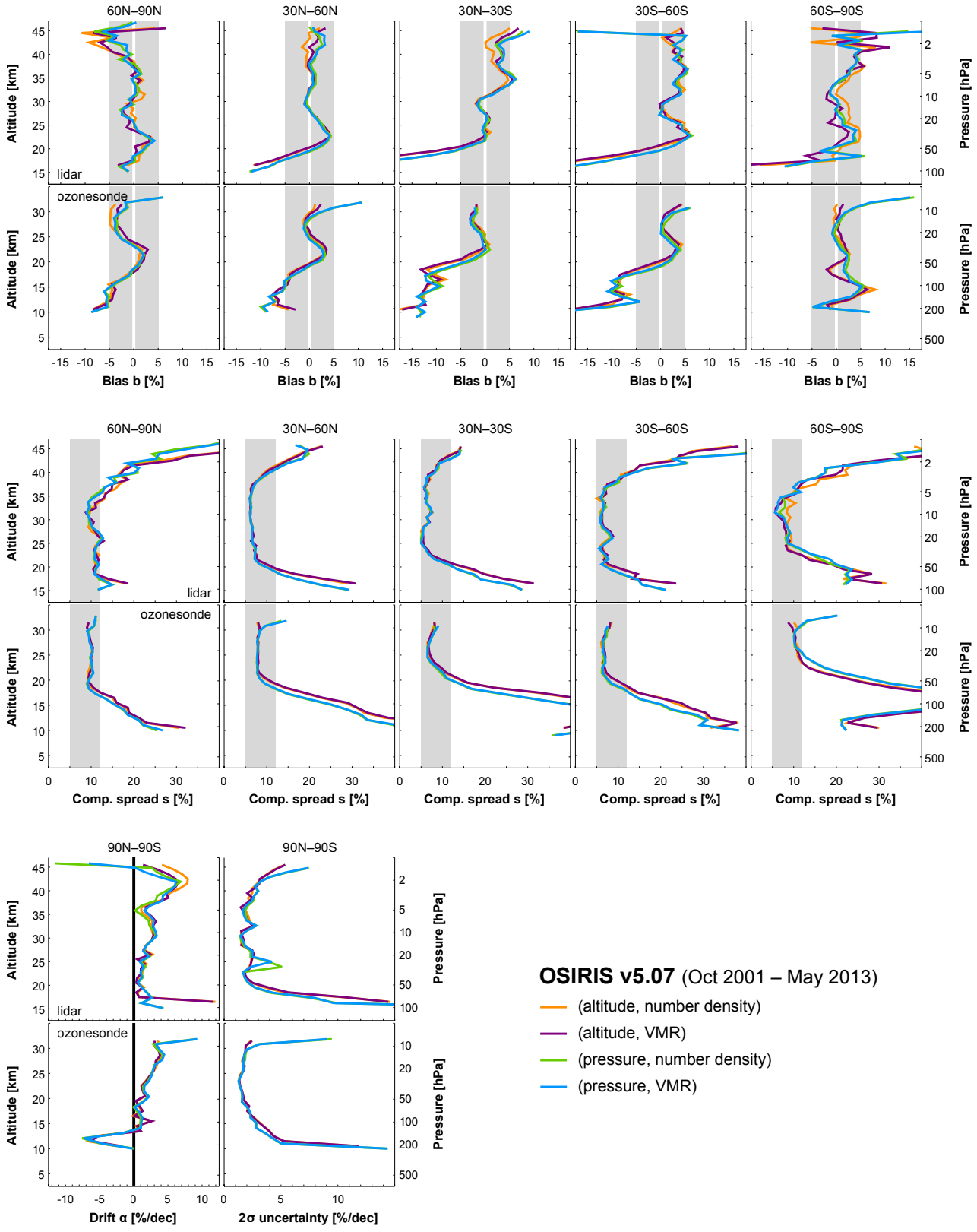

Figure S10. Same as Fig. S3, but for OSIRIS v5.07 (nominal representation: ozone number density on fixed altitude levels).

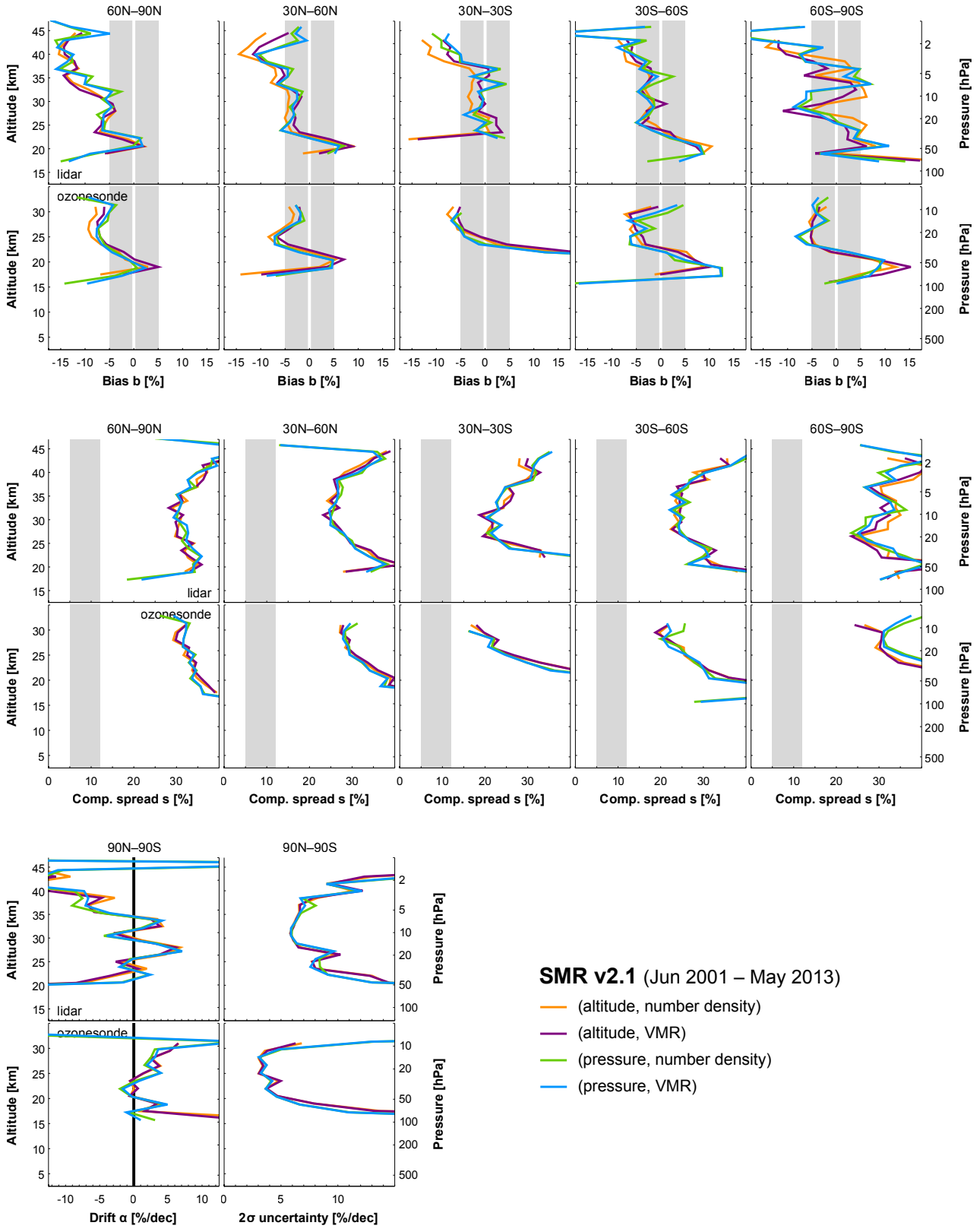

Figure S11. Same as Fig. S3, but for SMR v1.2 (nominal representation: ozone VMR on variable altitude levels).

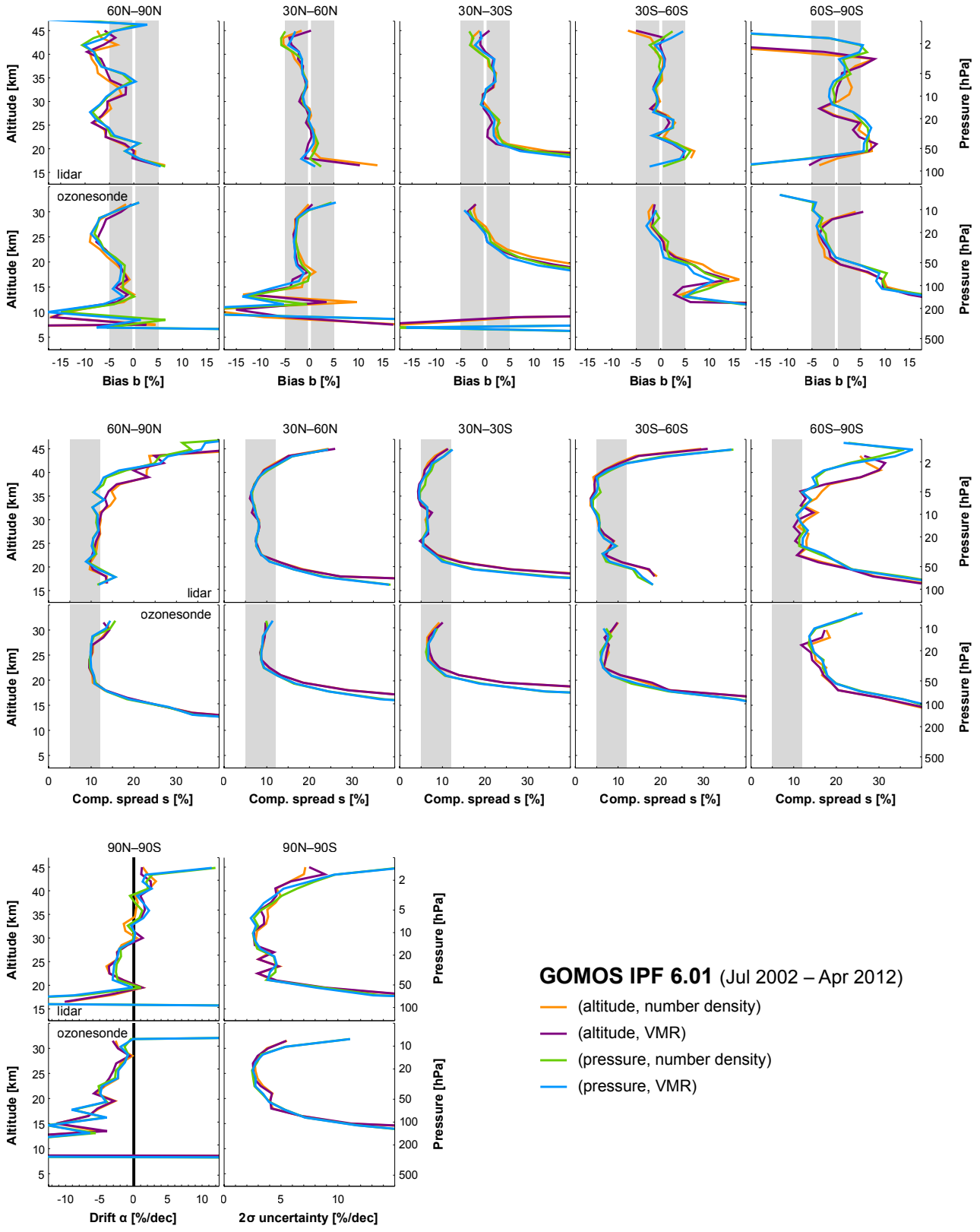

Figure S12. Same as Fig. S3, but for GOMOS IPF6.01 (nominal representation: ozone number density on variable altitude levels).

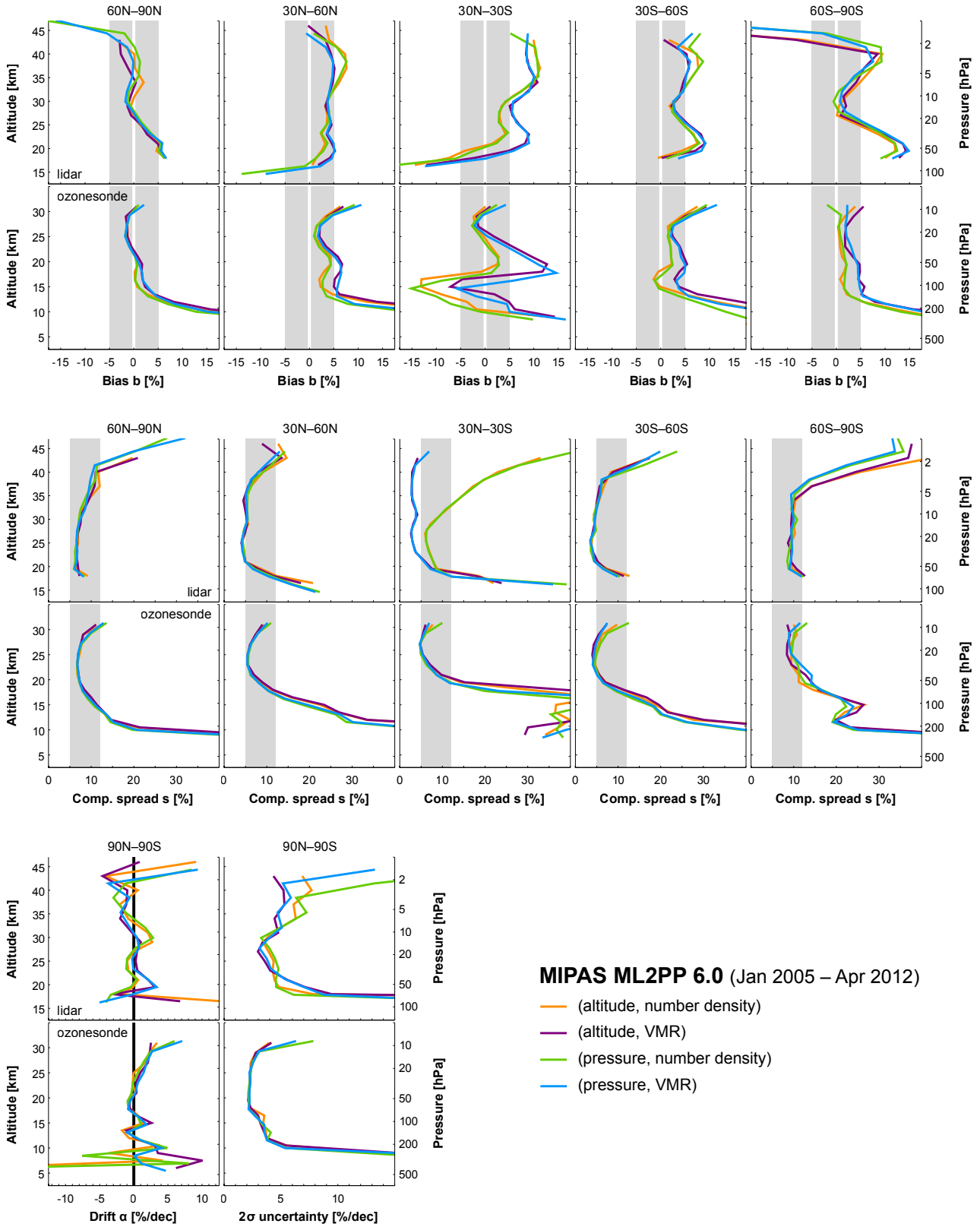

Figure S13. Same as Fig. S3, but for MIPAS ML2PP6.0 2005–2012 nominal observation data (nominal representation: ozone VMR on variable pressure levels). The large increase in the spread of tropical lidar comparisons is due to outliers in the averaging kernels. It is not seen when a triangular smoothing function is used.

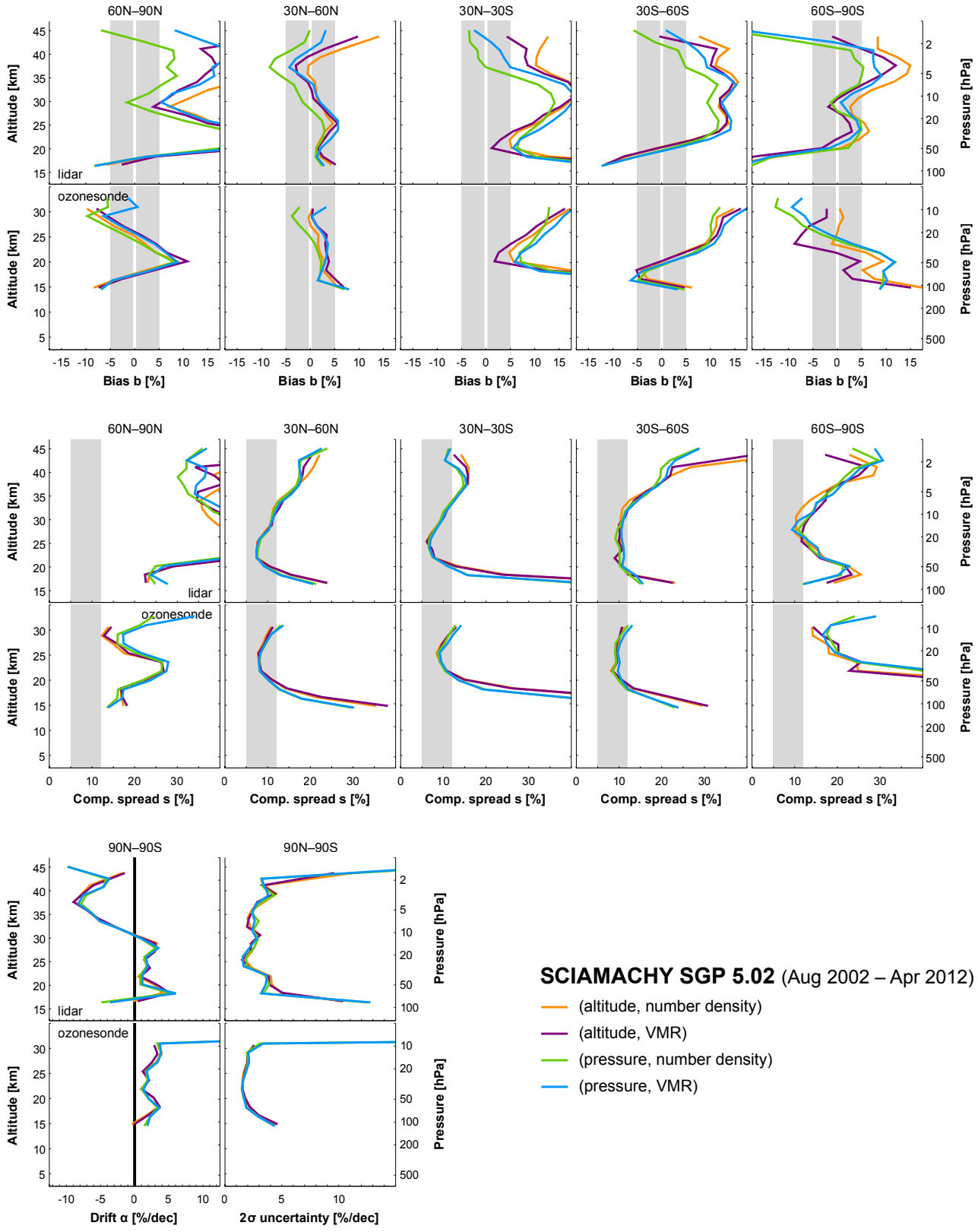

Figure S14. Same as Fig. S3, but for SCIAMACHY SGP5.02 (nominal representation: ozone number density on fixed altitude levels).

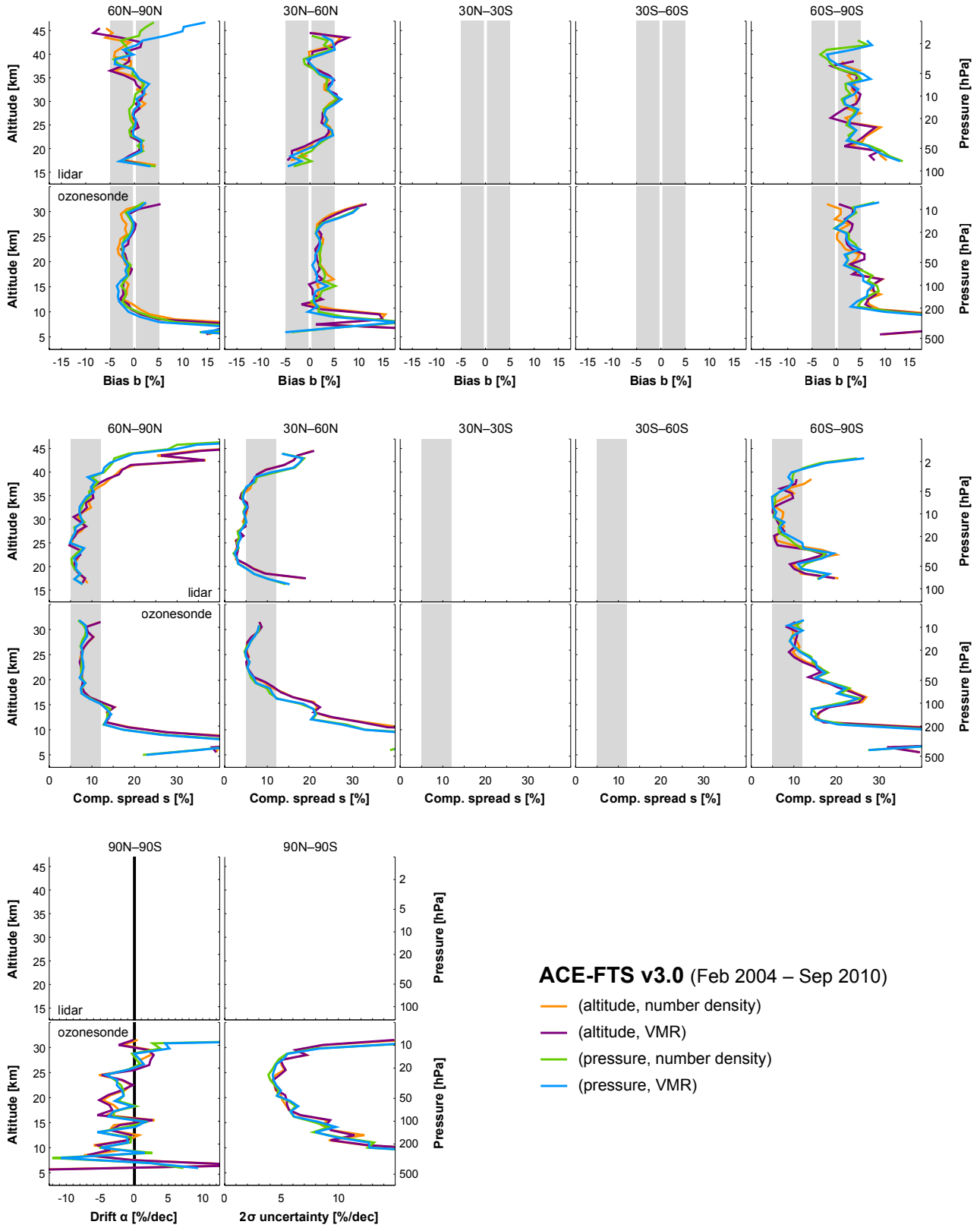

Figure S15. Same as Fig. S3, but for ACE-FTS v3.0 (nominal representation: ozone VMR on fixed altitude levels).

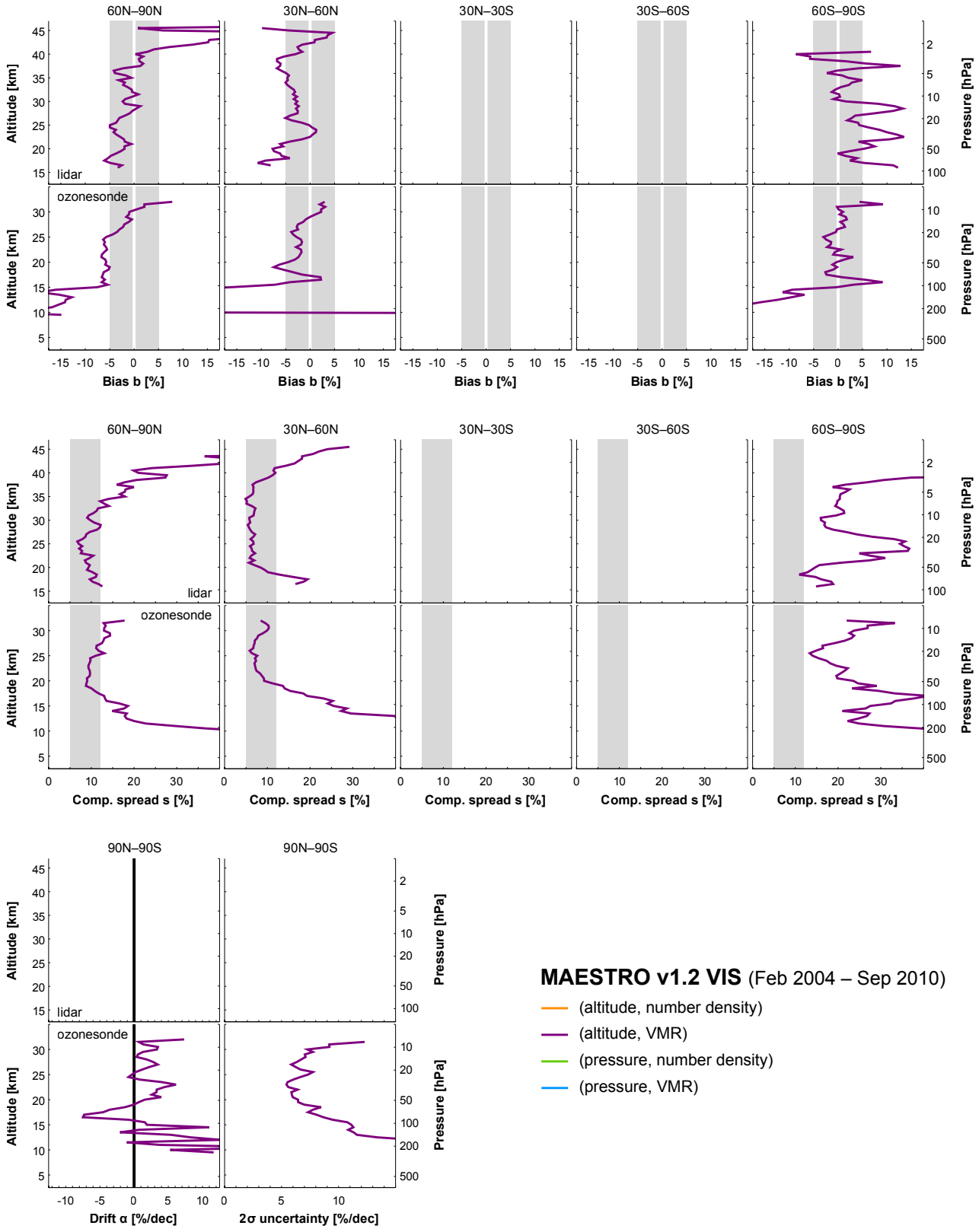

Figure S16. Same as Fig. S3, but for the nominal representation of MAESTRO v1.2 (ozone VMR on variable altitude levels). Non-native representations were not verified here since the required auxiliary data are identical to those of ACE-FTS v3.0. So we expect a similar representation-dependence as in Fig. S15.

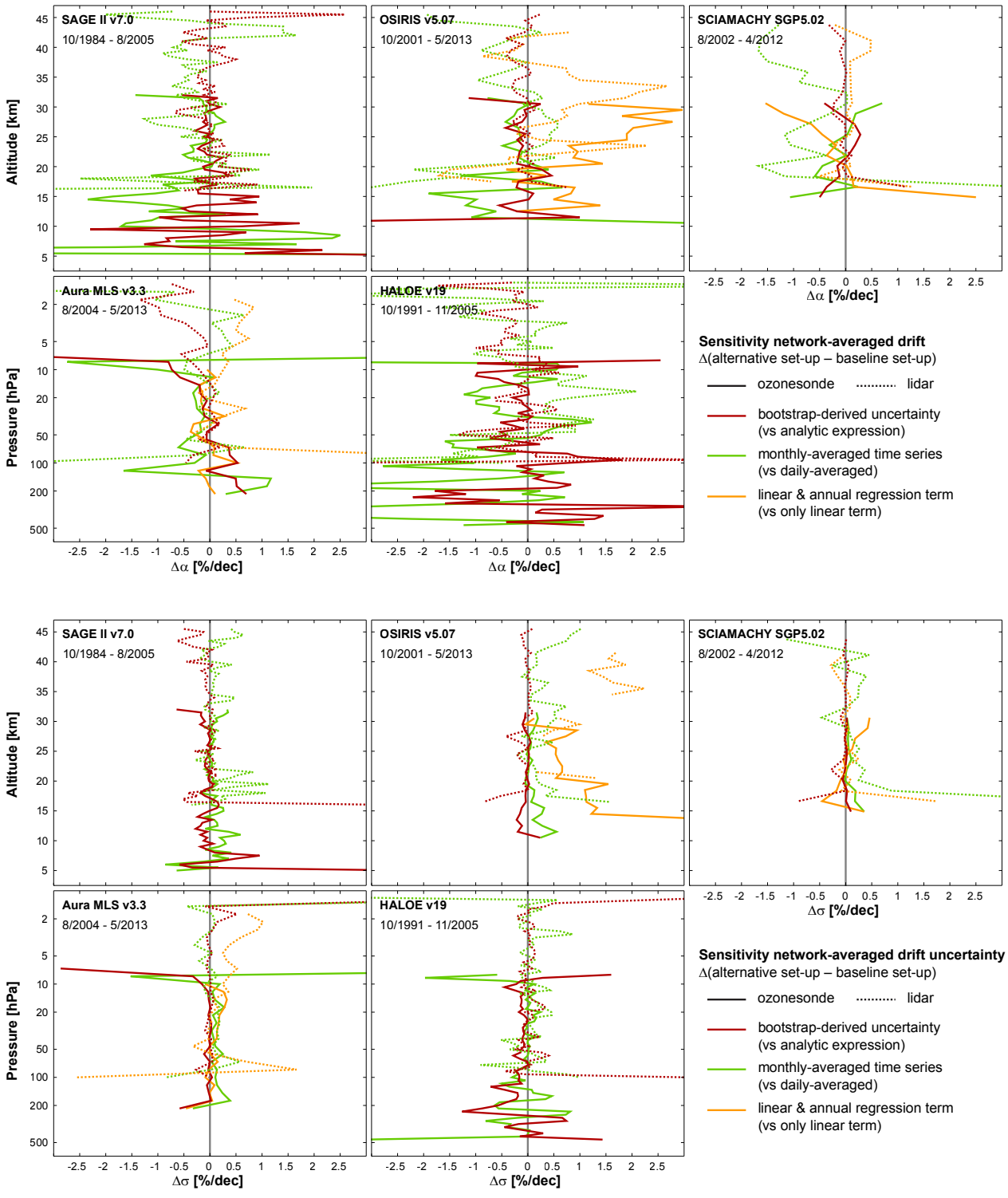

Figure S17. Sensitivity of ozonesonde and lidar network-averaged drift $\bar{\alpha}$ (top part) and one sigma uncertainty $\sigma^*_{\bar{\alpha}}$ (bottom part) to several regression analysis parameters (cfr. Sec. 4.1.3). Shown is the difference of an alternative minus the baseline regression set-up for five satellite records. Three alternatives are depicted: (a) uncertainties derived from bootstrap (red) instead of analytic expression, (b) regression of monthly-averaged comparison timeseries (green) instead of daily-averaged, and, (c) regression model with linear and annual terms (yellow) instead of only linear term. Larger differences appear for the latter set-up since only timeseries with a significant annual cycle are regressed, hence leading to a quite different sample of timeseries included in the network-averages.

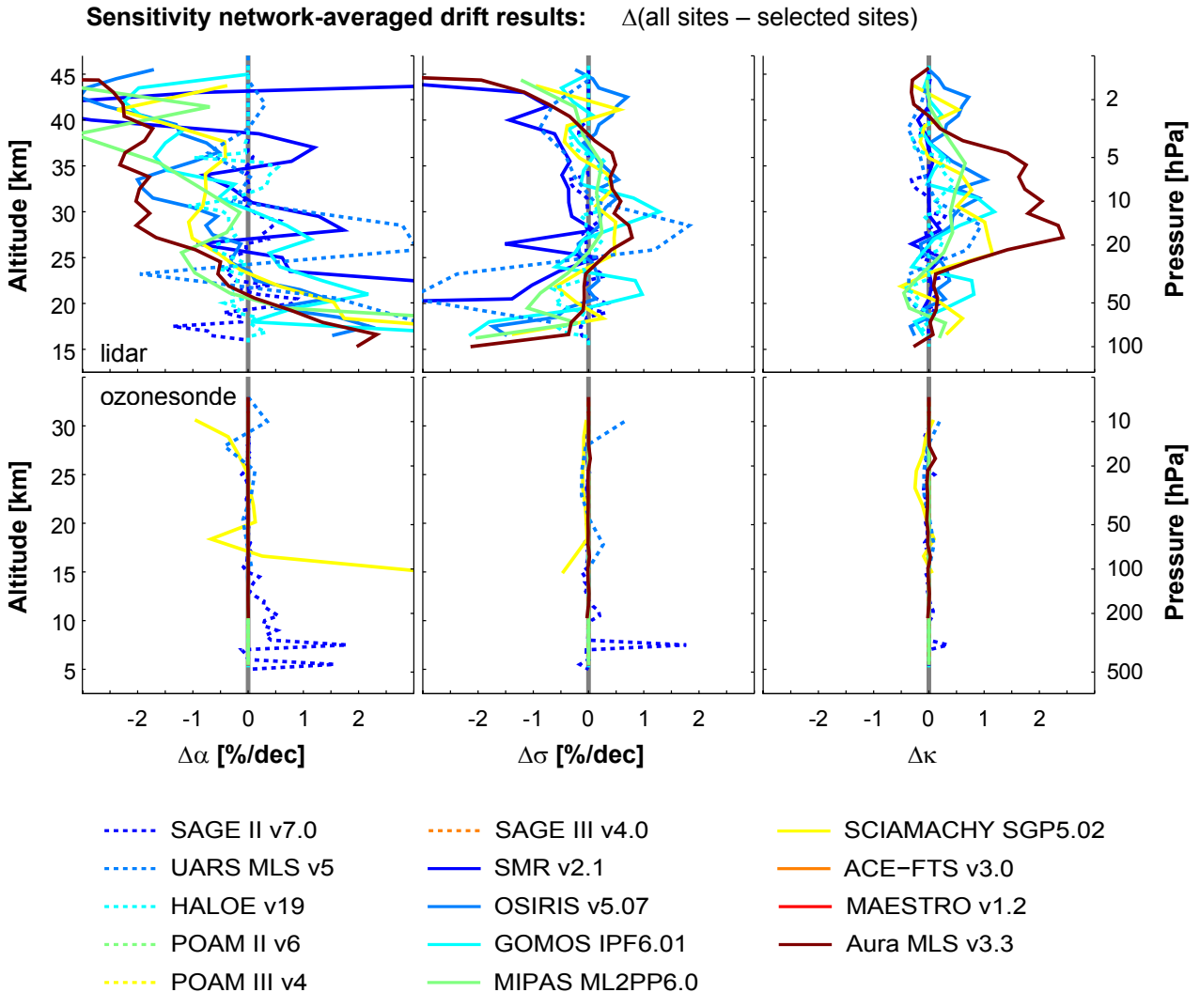

Figure S18. Sensitivity of ozonesonde (bottom) and lidar (top) network-averaged drift $\bar{\alpha}$, one sigma uncertainty $\sigma^*_{\bar{\alpha}}$ and κ-adjustment factor to the selection of ground sites (cfr. Sec. 4.2, Tables 1 and 2). Shown is the difference of the result obtained for the entire ground network minus the result for the selected sites, for the fourteen satellite records.

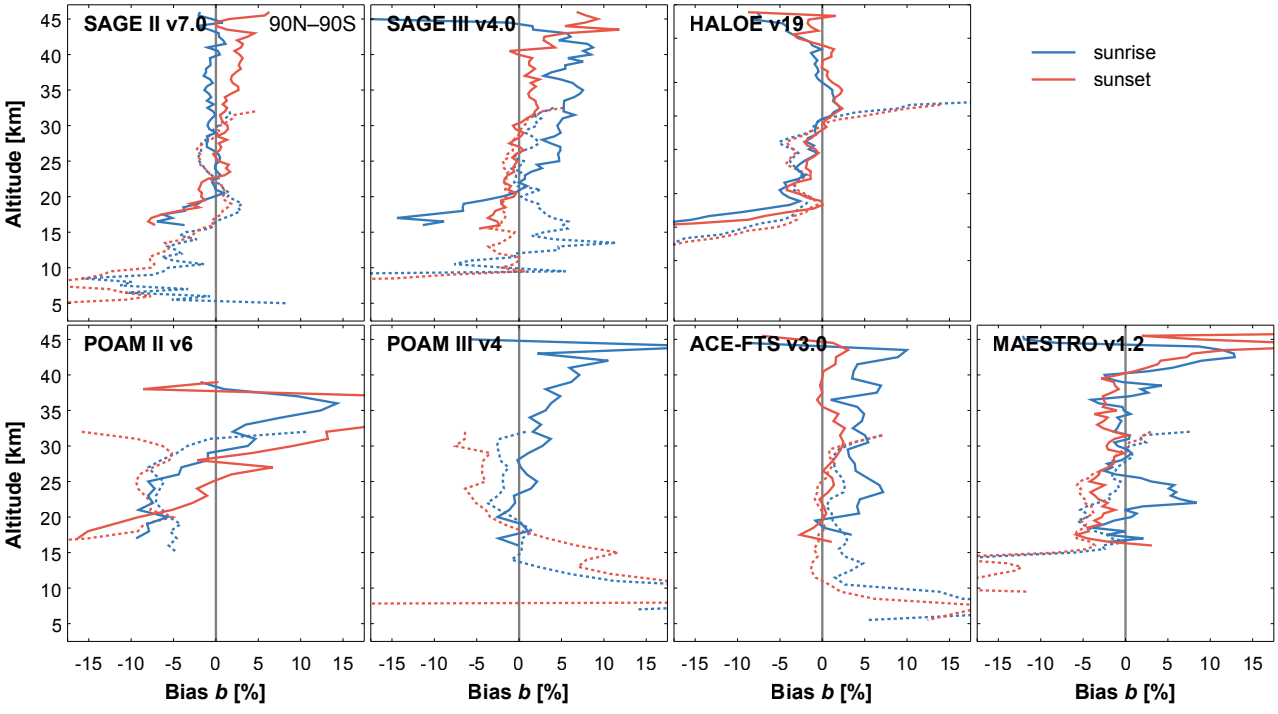

Figure S19. Bias relative to ozonesonde (dashed) and lidar (solid) of the sunrise and sunset measurements by seven solar occultation instruments. The bias is calculated over the entire ground-based network. The analysis is performed in the native profile representation of each satellite record.

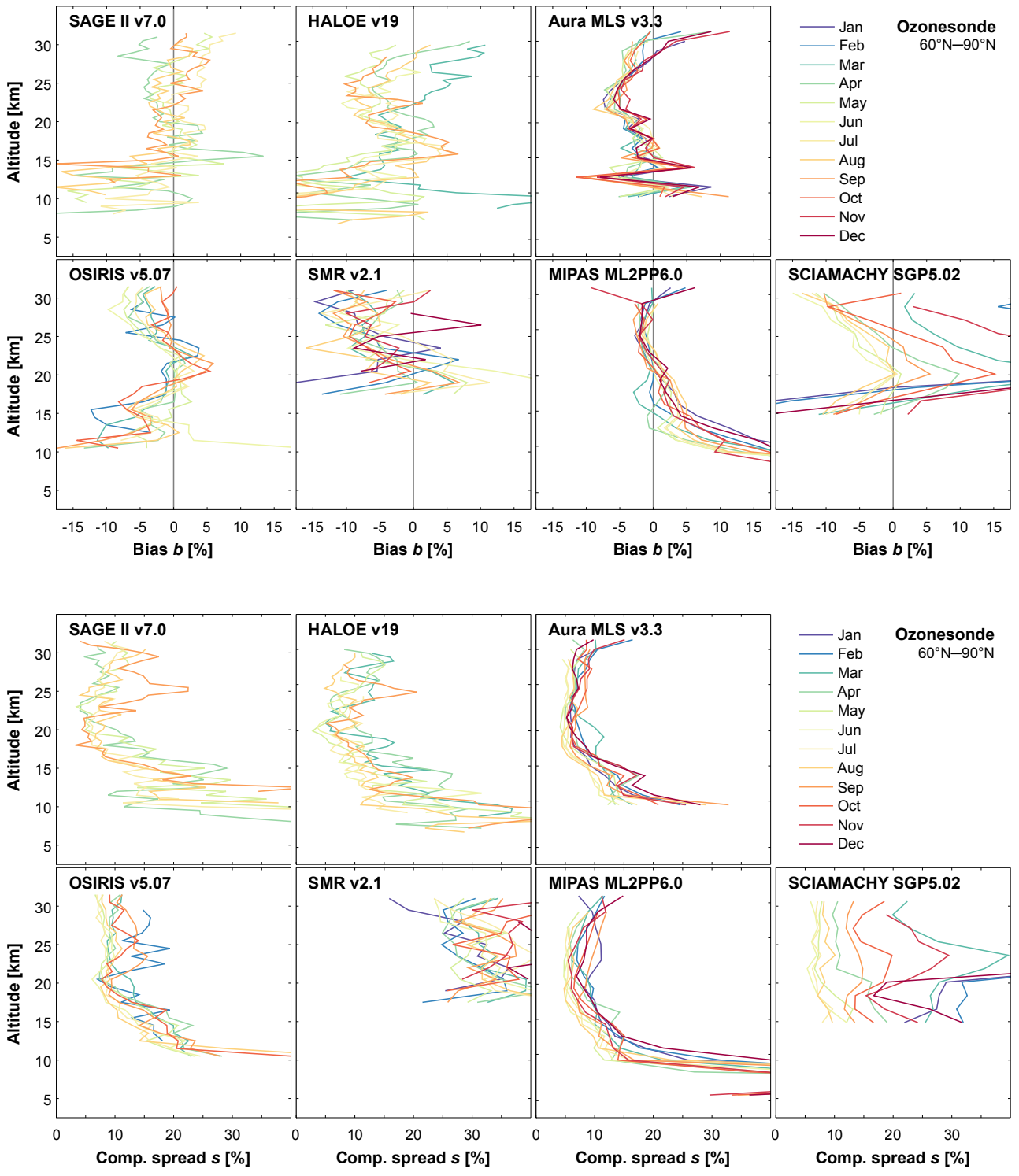

Figure S20. Dependence on month of the bias (top part) and spread (bottom) for comparisons of seven satellite records to Arctic ozonesonde data. The analysis is performed in the native profile representation of each satellite record.